

ESTUDO SOBRE A DIMENSIONALIDADE DO TESTE R1 – FORMA B

Fermino Fernandes Sisto*
Acácia Aparecida Angeli dos Santos#
Ana Paula Porto Noronha[¶]

RESUMO. O presente estudo objetivou utilizar o modelo de Rasch para analisar a possibilidade de unidimensionalidade do teste R1-Forma B, por meio da análise dos resíduos por componentes principais. Para tanto, foram investigados 745 estudantes que frequentavam um curso para jovens e adultos, caracterizando pessoas com defasagem na escolaridade ou em fase de escolarização tardia. Desses, 50,3% eram do sexo feminino e 48,3% do masculino, com idade variando entre 15 e 75 anos. Os resultados sugeriram, entre outros aspectos, que o Teste R1 não é unidimensional e que apenas uma pontuação global no teste não é suficiente para descrever as condutas cognitivas usadas para resolvê-lo. No entanto, a análise dos fatores 1 e 2 indicaram unidimensionalidade para ambos. Sugere-se que novos estudos sejam realizados com amostras diferentes, a fim de que sejam ampliadas as investigações aqui propostas.
Palavras-chave: inteligência, R1 – forma B, Teoria de Resposta ao Item.

STUDYING THE R-1 TEST - FORM B UNIDIMENSIONALITY

ABSTRACT. The possibility of unidimensionality of R1-Forma B test was analyzed by the Rasch model through the principal residual component analysis. Seven hundred and forty-five students attending a course for young people and adults were investigated. All subjects, 50.3% were females and 48.3% males, age bracket 15 - 75 years old, showed schooling discrepancy or delay. Besides other aspects, results suggested that the R1 Test is not unidimensional and that total score is insufficient to describe the cognitive behavior used to solve the issue. However, analyses of factors 1 and 2 indicated unidimensionality. Other studies with different samples are suggested so that investigations may be amplified and deepened.
Key words: Intelligence, R1 – forma B, Item Response Theory.

ESTUDIO SOBRE LA DIMENSIONALIDAD DEL TEST R1 – FORMA B

RESUMEN. El presente estudio tuvo el objetivo de utilizar el modelo de Rasch para analizar la posibilidad de unidimensionalidad del test R1-Forma B, por medio del análisis de los residuos por componentes principales. Para tanto, fueron investigados 745 estudiantes que frecuentaban un curso para jóvenes y adultos, caracterizando personas con retraso en la escolaridad o en fase de escolarización tardía. De esos, el 50,3% eran del sexo femenino y el 48,3% del masculino, con edad variando entre 15 y 75 años. Los resultados han sugerido, entre otros aspectos, que el Test R1 no es unidimensional y que tan sólo un puntaje global en el test no es suficiente para describir las conductas cognitivas usadas para resolverlo. Sin embargo, el análisis de los factores 1 y 2 indicaron unidimensionalidad para ambos. Se sugiere que nuevos estudios sean realizados con muestras distintas, a fin de que sean ampliadas las investigaciones propuestas aquí.
Palabras-clave: inteligencia, R1 – forma B, Teoría de Respuesta al Ítem.

No Brasil, há poucos trabalhos desenvolvidos com o objetivo de comprovar as qualidades psicométricas dos testes psicológicos, o que leva a um cenário desalentador quanto à avaliação psicológica,

* Doutor em Pedagogia. Professor da graduação e do Programa de Pós-graduação Stricto Sensu em Psicologia da Universidade São Francisco – Itatiba-SP.

Doutora em Psicologia Escolar e do Desenvolvimento Humano. Professora da graduação e do Programa de Pós-graduação Stricto Sensu em Psicologia da Universidade São Francisco – Itatiba-SP.

¶ Doutora em Psicologia: Ciência e Profissão. Professora da graduação e do Programa de Pós-graduação Stricto Sensu em Psicologia da Universidade São Francisco – Itatiba-SP.

principalmente devido ao fato de que são instrumentos freqüentemente usados no dia-a-dia do psicólogo, que por meio deles obtém informações para o processo de avaliação. Noronha (2001) avaliou 21 testes de inteligência publicados no Brasil, no que se refere aos coeficientes de validade, entre outros aspectos, verificando que a maioria deles continha as informações mínimas necessárias sobre as evidências de validade. Essa constatação já indicou um avanço em relação ao estudo feito por Sisto, Codenotti, Costa e Nascimento em 1979. Não obstante, as evidências de validade estão baseadas quase que exclusivamente em correlações com outros testes.

Embora no Brasil os estudos sobre os testes de inteligência sejam poucos quando se considera o cenário internacional, verifica-se que eles vêm sendo realizados há aproximadamente cem anos. Em um estudo de metaanálise, Schmidt e Hunter (1998) revisaram 85 anos de estudos de validade, demonstrando o valor de instrumentos bem-construídos.

Considera-se como marco dos estudos sobre a inteligência o modelo unidimensional de intercorrelações entre testes de Spearman (1904). Pode-se afirmar que o autor construiu um método estatístico para investigar os testes de inteligência, denominado de análise fatorial, com base na diferença tetrádica. Para o autor, dois fatores estariam sempre envolvidos na resolução de situações. O fator *g* foi considerado como uma capacidade básica que favoreceria o estabelecimento de relações e o pensamento abstrato (Ribeiro, 1998; Sternberg & Powell, 1983); e os fatores *s* facilitariam a realização de tarefas específicas que conteriam aspectos peculiares de uma dada situação-problema (Brody, 1992; Jensen, 1994). Em razão disso, sua teoria ficou conhecida como a teoria Bifatorial ou dos Dois Fatores. Em 1927, Spearman reconheceu que outros fatores poderiam interferir na manifestação de *g*. Assim, aceitou que aspectos como o cansaço e a motivação, entre outros, poderiam afetar os resultados obtidos durante a realização de um teste e, conseqüentemente, o valor de *g*. Além disso, apresentou algumas explicações para o fator *g*, defendendo que os raciocínios subjacentes a ele são a educação de relações e a educação de correlatos.

Thurstone (1931, 1938) pode ser considerado como um dos maiores críticos da proposta de Spearman, e como contraponto propôs uma teoria alternativa à dos Dois Fatores. Em seu modelo admitia a existência de sete capacidades mentais primárias, a saber, espacial (fator S), rapidez perceptual (fator P), numérica (fator N), compreensão verbal (fator V),

fluência verbal (fator W), memória (fator M) e o raciocínio indutivo (fator I). Foi com o emprego do método da análise fatorial que Thurstone chegou a esses fatores, e seus resultados indicaram a existência desses fatores, representando habilidades independentes (Almeida, 1988; Sternberg, 1986; Thurstone, 1931, 1938).

Posteriormente, Thurstone reviu a posição que assumira no que diz respeito à independência das habilidades que detectou. Em outros estudos ressaltou que havia uma alta correlação entre os fatores, admitindo haver muito mais semelhanças do que diferenças entre sua teoria e a de Spearman. Por essa ocasião passou a afirmar que um modelo hierárquico poderia descrever bem a forma como a inteligência se manifestava, englobando várias capacidades relacionadas entre si, que teriam como elemento comum o fator *g* (Thurstone & Thurstone, 1941).

O Teste R1- Forma B (Oliveira, 1973) pode ser incluído entre os vários testes construídos com base na proposta teórica do fator *g*, sendo um instrumento bastante usado quando se pretende avaliar pessoas com baixo nível de escolaridade ou com pouco conhecimento da língua portuguesa. Esse instrumento será analisado neste estudo, considerando a importância de pesquisas que visem populações como as referidas, uma vez que estas, assim como outras, necessitam ser submetidas à avaliação da inteligência (por exemplo, em exames psicotécnicos para obtenção de carteira de motorista, em processos seletivos de empresas, entre outros), o que justifica a busca por instrumentos com boa acuidade metrológica.

Considerando-se o pressuposto do teste R1-Forma B, é esperado que seja unidimensional, pois se propõe a avaliar o fator *g* da inteligência. No entanto, é necessário levar em consideração o estudo de Jensen (1998), quando afirma que a quantidade de variância de *g*, evidenciada em medidas padronizadas que estão sendo utilizadas, fica entre 64% e 75%, o que sugere a possibilidade da existência de outra dimensão. Vale lembrar que a unidimensionalidade é uma característica de testes que medem uma só dimensão psicológica, quando um único traço latente é avaliado nas respostas. A consistência interna não pode ser considerada como sinônimo de dimensionalidade, pois quando a unidimensionalidade é identificada há, necessariamente, alta consistência interna, mas o contrário não é verdadeiro. Ao lado disso, pode também haver testes com alta consistência interna que são multidimensionais.

Ao se estudar a evidência de validade de construto de uma dada medida psicológica, a análise de sua unidimensionalidade é de grande relevância,

visto que pode fornecer informações sobre a interpretação teórica dada ao escore obtido (APA, 1995). Cronbach e Meehl (1955) consideram a análise da unidimensionalidade como um estudo da consistência interna do teste, pois as pontuações dos itens que o compõem são relacionadas umas com as outras, verificando se todos os itens estão medindo o mesmo construto ou habilidade. Dessa forma, uma medida seria unifatorial quando a aferição de sua dimensionalidade demonstrasse que as respostas das pessoas para todos os itens do teste dependeriam de um único traço subjacente ou habilidade.

A identificação da unidimensionalidade de um teste tem implicações diretas na sua interpretação. No caso específico do R1- forma B, é possível supor que ele contenha dois grupos de itens, isto é, itens relacionados à educação de relação e itens relacionados à educação de correlatos. Assim, sabe-se que a pontuação de um indivíduo é resultante de uma combinação de ambos os tipos de educação. Vale considerar um exemplo no qual duas pessoas obtêm 20 pontos: uma delas pode apresentar 15 pontos em educação de relações e apenas 5 pontos em educação de correlatos, e a segunda poderá fazer os mesmos 20 pontos, sendo 15 deles em educação de correlatos e um escore menor (5 pontos, por exemplo) em educação de relações.

Vê-se, destarte, que apenas a aferição da pontuação geral do sujeito no teste não traz em si a informação sobre a combinação do raciocínio por ele utilizado, que pode ser totalmente ignorada. Pelo exemplo citado verifica-se que, com base na pontuação total obtida no teste R1-forma B, é possível supor que seja difícil realizar uma interpretação teórica adequada caso as eduções componham fatores diferentes, ou seja, não sejam unidimensionais.

O autor do teste (Oliveira, 1973) parte do pressuposto que o instrumento avalia o fator *g*. Também o manual do teste R1 (Alves, 2002), que é uma forma paralela do R1-Forma B, apresenta uma fundamentação que permite a interpretação de que deveria ser unidimensional, pois estaria medindo o fator *g*. O estudo de validade convergente, apresentado no manual, é feito por análise de correlação com o Teste de Matrizes Progressivas do Raven, cujo manual (Cunha, 2001; Raven, Raven & Court, 1991) também afirma que se trata de uma medida de *g*. Em decorrência, reforça a proposta do autor de que se trata também de um teste que avaliaria o fator *g*. Todos eles propõem uma avaliação pela pontuação total do teste, o que pressupõe que eles sejam medidas unidimensionais.

Não obstante, algumas particularidades em sua construção e correção facilitam também a interpretação de que sua estrutura poderia não ser unidimensional. No manual do R1 (Alves, 2002) há a descrição dos vários tipos de raciocínios subjacentes em cada item. Outra informação aponta para grandes diferenças entre pessoas com níveis de escolaridade diferentes, pois amostras compostas por sujeitos com escolaridade mais elevada apresentaram diferenças significativas em relação às de menor escolaridade. Tais características indicam a possibilidade de que o teste contenha itens dependentes de aprendizagem e não estaria medindo apenas o fator *g*, portanto não seria unidimensional.

Alguns métodos estão disponíveis para análise da dimensionalidade de uma escala. Os estudos de validação 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; Ender, Parker & Summerfeldt, 1998). No entanto, a análise fatorial, quando aplicada a variáveis binárias, como é o caso do Teste R1 e R1-Forma B, produz quase sempre muitos fatores, porque de fato é uma análise multifatorial e, em geral, alguns dos fatores são diferenciados em razão da dificuldade do item, chegando mesmo a extrair fatores artificiais (Green, Lissitz, & Mulaik, 1977; Hattie, 1985; McDonald & Ahlawat, 1974). Vale ressaltar que a média e o desvio-padrão não podem causar fatores artificiais, principalmente porque a análise fatorial considera pontuações padronizadas, nas quais os efeitos da média e dos desvios-padrão são parcialmente retirados. Acrescenta-se a isso o fato de que os modelos de análise fatorial podem não ser válidos nesses casos, porque foram baseados no pressuposto de que o traço latente subjacente apresenta sempre uma distribuição dentro da curva normal e não foram desenvolvidos para trabalhar com variáveis binárias (Bartholomew, 1980; Christoffersson, 1975).

Sem o pressuposto de que o traço latente tenha uma distribuição normal (Allen & Yen, 1979; Lord, 1953; Van der Linden & Hambleton, 1997), os modelos de Teoria de Resposta ao Item (TRI) foram propostos especialmente para a análise de variáveis binárias. Dentre os modelos de TRI, o de Rasch (1960) foi escolhido para este estudo por algumas razões técnicas, principalmente pelo fato de 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).

Outro modelo que poderia ser cogitado para esse tipo de dado é o modelo logístico de dois parâmetros. Embora esse modelo forneça muitos testes estatísticos, há nele carência de fundamentos matemáticos rigorosos, o que não é o caso de algumas estatísticas do modelo Rasch. Deve-se acrescentar também que o modelo logístico de dois parâmetros possui teste estatístico pouco desenvolvido, baseado em frequências bivariadas, exatamente nos aspectos em que as violações da unidimensionalidade são mais prováveis (van den Wollenberg, 1982).

Destaca-se que, quando se trata de estudar resultados de medidas psicométricas, três utilidades são valorizadas para a escolha da análise pelo modelo de Rasch. Uma delas é fornecer uma seqüência hierárquica de itens de acordo com a dificuldade; outra é avaliar o ajuste de itens e determinar quais deles pertencem ao mesmo traço latente; e, finalmente, a terceira, revelar grupos de itens de mesmo nível de dificuldade.

Vale destacar também que há um critério bastante claro de unidimensionalidade para verificar se os dados se ajustam ao modelo, impossibilitando a artificialidade. Assim, no modelo de Rasch (1960) a verificação do ajuste dos dados ao modelo está fundamentada na suposição de que a pessoa, ao responder a um item do teste, possui alguma habilidade. Tal habilidade é passível de ser representada por um valor numérico ou pontuação, que indica seu lugar em uma dada escala. Pressupõe-se que para cada nível de habilidade existe uma probabilidade - previamente determinada pelo modelo - de que as pessoas com esse nível de habilidade dêem a resposta correta para o item; isto é, para pessoas de baixa habilidade a probabilidade esperada é pequena, enquanto para as de alta habilidade, a probabilidade é alta.

Nesse sentido, é possível afirmar que 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. Vale lembrar que a fórmula na qual o modelo está baseado admite que essas características são estáveis para todos os sujeitos e itens do teste, o que caracteriza a referida unidimensionalidade. Assim, a pessoa usa a mesma habilidade para responder a todos os itens, sendo que a dificuldade relativa do item permanece inalterável para todas as pessoas que a ele respondem, ou seja, a habilidade do sujeito permanece constante por todo o teste.

Uma das vantagens do uso do modelo é que os itens podem ser organizados por ordem de dificuldade, de forma que essa ordem independa das habilidades das

pessoas, bem como de características pessoais - como inteligência, motivação, interesse - ou ambientais, como barulho ou calor, não sendo modificada a ordem encontrada. Há também outro aspecto que merece destaque, quando se considera o processo de avaliação. Trata-se da chamada "suficiência da pontuação total" (Lord, 1980; Smith & Miao, 1994), considerada como "medida suficiente" para a habilidade subjacente, visto que ela contém toda a habilidade das pessoas da amostra na escala em que se está avaliando. Assim, pode-se afirmar, por um lado, que nenhuma informação relevante se perde ao se considerar a pontuação total e, por outro, que a pontuação total não contém outras informações além da habilidade em questão.

Considerando-se o modelo de Rasch, se o Teste R1- Forma B apresentar um bom índice de ajuste, a pontuação total da pessoa será um resumo da habilidade subjacente ao teste, sendo possível interpretá-la. O não-ajuste ao modelo poderia ser interpretado como indicador de que a pontuação total não é uma medida suficiente para indicar a quantidade da habilidade subjacente e, por decorrência, haverá dificuldade para se interpretar seu resultado à luz do arcabouço teórico sob o qual foi proposto, refletindo uma combinação das habilidades envolvidas, cuja forma de aglutinação se desconhece. A implicação conseqüente envolveria a necessidade de se mudar a forma de pontuar o teste R1-forma B, visto que, na proposta original de Oliveira (1973), sua avaliação é feita pela pontuação total. Tal cuidado seria necessário para que não fossem desconsideradas dimensões relevantes que, na realidade, poderiam estar sendo medidas pelo teste. Uma alternativa seria buscar encontrar uma organização dos itens (com retirada ou acréscimo dos atualmente existentes) de forma que se ajustassem à proposta de unidimensionalidade; outra seria utilizar mais de uma medida, em razão de quantos fatores forem necessários para explicar os itens.

Ao lado disso, é necessário considerar que o Manual do R1-Forma B (Sisto, Santos & Noronha, 2004) propõe dois fatores para interpretar os itens desse teste. No entanto, para chegar a essa conclusão os autores trabalharam com análise fatorial por eixos principais. Tecnicamente falando-se, trata-se de um modelo multifatorial que foi usado para dados dicotômicos. Não se pode deixar de lado o fato de que os dois fatores do R1-Forma B mostraram correlação positiva de 0,49 e, com base nessa correlação, pode-se supor que há um fator subjacente a esses dados. Tais aspectos

sugerem a necessidade de mais análises, com vistas a verificar se são necessários dois fatores ou não.

Das várias análises possíveis para verificar esse ajuste (van den Wollenberg, 1982; Glas, 1988), optou-se pela análise de componentes principais de resíduos, proposta por Linacre em 1996. Nessa análise pressupõe-se que o modelo Rasch extrai a melhor dimensão possível dos dados obtidos, não podendo ser negada a possibilidade de que cada item contere ainda informações peculiares e os distintos itens poderão apresentá-las em diferentes graus. Conseqüentemente, essas peculiaridades poderão indicar eventuais contradições em relação à unidimensionalidade intrínseca do item. Para isso, analisa-se a diferença entre o que o modelo prediz e o que é, de fato, observado. Essa diferença é denominada de resíduos e pode ser avaliada pela análise dos componentes principais, por meio da qual indicações de estruturas secundárias ou subdimensões dos dados poderão ser encontradas.

Dentro desse contexto, o presente estudo teve como objetivo utilizar o modelo de Rasch para analisar a possibilidade de unidimensionalidade do teste R1-Forma B, por meio da análise dos resíduos por

componentes principais, o que parece oportuno para uma análise psicométrica do referido instrumento. Ao mesmo tempo, avaliar a proposta de dois fatores para representar os mecanismos usados para responder ao teste, conforme indicado pelos autores do Manual (Sisto, Santos & Noronha, 2004).

MÉTODO

Participantes

Foram investigados 745 estudantes que freqüentavam o curso de Educação de Jovens e Adultos (EJA), caracterizando pessoas com defasagem na escolaridade ou em fase de escolarização tardia. Do total de estudantes, 50,3% eram do sexo feminino e 48,3% do sexo masculino. As idades variaram de 15 a 75 anos, situando-se mais acentuadamente na faixa entre 17 e 39 anos, responsável por 74,1% dos sujeitos. A média foi de 32,11 (mediana de 30 e moda de 18) e o desvio-padrão, de 12,16. A seguir são apresentadas na Tabela 1 as freqüências de participantes por faixa etária, nível de escolaridade e gênero. Após a aprovação do Comitê de Ética e a autorização das escolas, deu-se início à coleta de dados.

Tabela 1. Freqüência de Participantes por Faixa Etária, Nível de Escolaridade e Sexo.

Faixa etária	Não informado	1		2		3		4		Total		
		I Grau	II Grau	I Grau	II Grau	I Grau	II Grau	I Grau	II Grau			
Sexo	F	2	14	53	71	18	59	42	32	37	13	341
	M	6	5	52	102	38	61	31	29	8	6	338
Não informado		10		16		21		14		5		66
Subtotal		8	19	105	173	56	120	73	61	45	19	679
Total		37		294		197		148		69		745

Instrumento

O material utilizado no presente estudo foi o Teste R-1-Forma B, construído em 1973 por Oliveira. O instrumento é composto por 40 itens, apresentados em um caderno, havendo para cada questão de 6 a 8 possibilidades de resposta. Os itens iniciais são constituídos de algumas figuras de objetos concretos, a fim de facilitar a compreensão da tarefa, e os demais são figuras geométricas.

Procedimentos

A aplicação do instrumento foi coletiva, em grupos de 15 a 20 sujeitos, havendo em cada sala pelo menos um aplicador e um auxiliar. O tempo total de aplicação do teste foi de 50 minutos.

Primeiramente foi feita a análise dos componentes principais dos resíduos. Dessa análise foram fornecidas as proporções do *Yardstick* e do ruído, como dados

descritivos, além das unidades explicadas em razão do total de unidades, que foi usado como critério neste estudo. Um teste com extensão razoável precisaria ter pelo menos três itens para afirmar a presença de uma dimensão secundária; caso tenha menos, é provável que ele seja unidimensional. Ao lado disso, os estudos de simulação de Smith e Miao (1994) indicaram que, nessas condições, eigenvalues menores que 1,4 podem ser considerados aleatórios. Assim que os resultados se encaixaram nesses critérios sugestivos de mais de uma dimensão, a análise teve continuidade, 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 as pontuações de ambos os fatores foram correlacionadas.

Para interpretar as correlações encontradas entre os fatores extraídos dos resíduos, foi assumida a hipótese 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 menos 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.

RESULTADOS

As estatísticas descritivas dos dados mostraram que a média foi de 21,14 ($DP=8,59$) para valores mínimo e máximo de 1 e 36, respectivamente, no teste que contém 40 itens. Foi analisada também a consistência interna (alfa de *Cronbach*), cujo valor obtido foi de 0,93.

Em seguida, investigou-se a dimensionalidade do instrumento, analisando-se as correlações dos resíduos padronizados. Para isso empregou-se a análise fatorial por componentes principais. A Tabela 2 fornece os principais resultados dessa análise para os itens selecionados em razão da maior saturação, positiva e negativa, indicando, assim, suas características descritivas. A proporção do fator no *Yardstick* foi de 70,3/1 e a proporção do ruído, de 5,5/1. O Fator 1 explicou 3,2 de 40 unidades de variância residual, o que indica a possibilidade de que, extraída a variância da escala produzida pelo modelo Rasch, ainda ficaram resíduos suficientes para configurar um outro fator.

Tabela 2. Itens com Carga Fatorial Superior a 0,24, *Measure*, *Infit* e *Outfit* do Fator 1, Extraído por Análise de Componentes Principais de Correlações de Resíduos Padronizados.

Itens	cargas	medida	Infit	Outfit
12	0,61	-1,82	0,75	0,55
9	0,59	-1,71	0,76	0,54
11	0,59	-1,67	0,71	0,51
6	0,48	-2,18	0,80	0,48
8	0,46	-0,97	0,86	0,72
13	0,43	-1,52	0,87	0,75
10	0,35	-1,26	0,89	0,96
18	0,34	-0,87	0,78	0,65
19	0,28	-1,37	0,71	0,55
40	-0,38	3,06	1,32	6,77
36	-0,31	3,32	1,29	4,98
33	-0,31	1,73	0,96	1,30
30	-0,31	-0,73	0,92	0,83
38	-0,30	2,79	1,07	2,87
29	-0,29	0,37	1,17	1,52
37	-0,26	3,06	1,15	3,16
27	-0,25	1,05	1,18	1,30
17	-0,25	1,02	1,15	2,48

Os itens com cargas positivas selecionados foram 12, 9, 11, 6, 8 e 13 e os com cargas negativas foram os itens

40, 36, 33, 30, 38 e 29. Os dois conjuntos de itens foram submetidos ao modelo de Rasch, para novo cálculo das pontuações dos sujeitos. Com as novas pontuações de cada sujeito, nos dois conjuntos, foram calculadas as correlações, e o coeficiente encontrado foi de $r=0,21$. Esse coeficiente foi corrigido pela fórmula de profecia Spearman-Brown, tendo por base a extensão de 20 itens, o que evidenciou um valor de $r=0,37$. Este coeficiente, muito inferior a 0,51, sugere que os dois conjuntos, cujos resíduos estavam mais discrepantes, possuem uma parte não desprezível da variância não explicada pela comunalidade entre eles. Dessa forma, essa análise sugere a necessidade de outro fator para explicar a parte da variância que a dimensão captada como preponderante pelo modelo de Rasch não explica. São fortes os indícios de que o Teste R1 não é unidimensional.

Em razão dos estudos feitos usando-se os resíduos fatorialmente e o modelo de Rasch, chegou-se à conclusão de que apenas uma pontuação global no Teste R1 Forma B não é suficiente para descrever as condutas cognitivas usadas para resolvê-lo. De certa forma, esse resultado corrobora a decisão de usar dois fatores para aglutinar os itens do referido teste, conforme o Manual (Sisto, Santos & Noronha, 2004), apesar de não ter sido essa a proposta inicial do autor do teste (Oliveira, 1973). Em razão disso, foi analisado cada um dos fatores extraídos pela análise fatorial de eixos principais, segundo a proposta do Manual.

A mesma análise de componentes principais dos resíduos foi feita com o Fator 1 do Teste R1-Forma B, cujos principais resultados encontram-se na Tabela 3, fornecendo informações descritivas sobre suas características. A proporção do fator no *Yardstick* foi de 83,0/1 e a proporção do ruído foi de 9,3/1. Novamente, este estudo indicou a possibilidade de ter ficado resíduo suficiente para configurar um outro fator, pois o Fator 1 explicou 2,2 de 20 unidades de variância residuais.

Tabela 3. Itens com Carga Fatorial Superior a 0,24, *Measure*, *Infit* e *Outfit* do Fator 1, Extraído por Análise de Componentes Principais de Correlações de Resíduos Padronizados.

Itens	cargas	medidas	Infit	Outfit
11	0,56	-0,24	0,74	0,44
12	0,55	-0,42	0,75	0,41
9	0,51	-0,28	0,77	0,41
16	0,43	-0,83	0,85	0,49
8	0,37	0,61	0,87	0,61
13	0,34	-0,05	0,93	0,60
20	-0,39	1,17	1,11	1,11
3	-0,38	0,03	1,15	0,82
15	-0,35	1,35	1,25	1,18
36	-0,31	5,62	1,32	9,90
7	-0,29	-1,03	1,05	1,14

Foram selecionados os itens 11, 12 e 9 com cargas positivas e os itens 20, 3 e 15 com cargas negativas, em razão de serem os itens com as maiores saturações. O mesmo procedimento descrito para o teste como um todo foi realizado e o coeficiente de correlação encontrado foi de $r=0,54$. Esse coeficiente foi corrigido pela fórmula de profecia Spearman-Brown, tendo por base a extensão de 10 itens, e forneceu um valor de $r=0,80$. Esse coeficiente, muito superior a 0,51, sugere que os dois conjuntos, cujos resíduos estavam mais discrepantes, produziram uma variância suficiente para ser descartada a possibilidade de uma outra dimensão. Assim, essa análise sugere a unidimensionalidade do Fator 1 do Teste R1- Forma B. A escala obtida nesse fator é suficiente para explicar os mecanismos subjacentes a ela.

Tabela 4. Itens com Carga Fatorial Superior a 0,24, *Measure*, *Infit* e *Outfit* do Fator 2, Extraído por Análise de Componentes Principais de Correlações de Resíduos Padronizados.

Itens	cargas	medida	<i>Infit</i>	<i>Outfit</i>
14	0,51	-0,65	0,95	1,05
24	0,43	-1,72	0,85	0,80
17	0,42	-0,46	1,12	1,39
25	0,37	-0,27	0,92	0,85
23	0,25	0,03	0,95	1,09
29	-0,44	-1,11	1,03	1,09
40	-0,43	1,59	1,29	2,30
27	-0,41	-0,44	1,12	1,14
30	-0,32	-2,20	0,74	0,61
37	-0,30	1,59	1,11	1,64

Os resultados da análise de componentes principais dos resíduos do Fator 2 do Teste R1-Forma B estão na Tabela 4. A proporção do fator no *Yardstick* foi de 12,3/1 e a proporção do ruído no *Yardstick* foi de 1,1/1. Os dados indicaram a possibilidade de que ficou resíduo suficiente para configurar um outro fator, pois o Fator 2 explicou 1,7 de 20 unidades de variância residuais.

Os itens com cargas positivas selecionados foram o 14, o 24 e o 17, e os com cargas negativas, os itens 29, 40 e 27, seguindo-se o mesmo critério. Aplicado o procedimento já comentado, o coeficiente de correlação encontrado foi de $r=0,24$, que corrigido pela fórmula de profecia Spearman-Brown, tendo por base a extensão de 10 itens, resultou em um valor de $r=0,51$. Esse coeficiente, igual a 0,51, sugere que os dois conjuntos, cujos resíduos estavam mais discrepantes, não possuem variância suficiente para compor um outro fator.

Assim, esse resultado admite a unidimensionalidade do Fator 2 do Teste R1- Forma B. A escala obtida nesse fator é suficiente para explicar os mecanismos subjacentes a ela.

DISCUSSÃO

Tendo em vista o objetivo de avaliar a dimensionalidade do Teste de Inteligência R1-Forma B, vale lembrar a importância do desenvolvimento de estudos para avaliar a qualidade psicométrica de testes psicológicos usados pelo psicólogo em sua prática profissional. Outros trabalhos têm ressaltado a carência de pesquisas nesse sentido (Noronha, 2001, por exemplo).

Muitos testes de inteligência foram elaborados com base na proposta teórica de Spearman (1904), pressupondo a existência do fator *g*, subjacente às atividades realizadas pelos sujeitos. Assim, o Teste de Matrizes Progressivas de Raven (Cunha, 2001) e o G36 (Boccalandro, 2003) estão entre os vários instrumentos bastante conhecidos e utilizados no Brasil que foram construídos para medir o fator *g*. Por ser um instrumento proposto para avaliação da inteligência de pessoas com baixo nível de escolaridade ou com pouco conhecimento da língua portuguesa é preciso que se tome um cuidado redobrado, visto que facilmente o resultado pode ficar comprometido em decorrência das características da população à qual se destina.

Neste estudo a unidimensionalidade do teste de R1-Forma B foi testada pelo modelo de Rasch, estudando-se o teste como um todo e cada fator separadamente. Oliveira (1973) o criou sob a inspiração do fator *g*, tal como o Teste de Matrizes Progressivas de Raven e o G36, e a indicação primeira foi que as pessoas deveriam ser avaliadas por uma única pontuação (a soma de todos os itens corretos), denotando a pressuposição de que se trataria de uma medida unidimensional. Contudo, o resultado das análises de todos os itens do teste não mostrou indícios que permitam afirmar sua unidimensionalidade, o que, de certo modo, não fornece base para a proposta original do autor do teste, que afirmou se tratar de uma medida unidimensional do fator *g*.

Assim, em relação ao teste como um todo, o sistema que o modelo Rasch conseguiu produzir para colocar os itens na escala não usou de forma mais ou menos proporcional a variância de todos os itens. Em outros termos, as análises sugeriram que a unidimensionalidade do R1-Forma B carece de sustentação e a pontuação total, como foi proposta pelo autor do instrumento para sua avaliação, pode

não ser uma medida das habilidades envolvidas suficiente para as pessoas responderem ao instrumento. Em contraposição, no entanto, confirmam a proposta de dois fatores feita pelos autores do Manual (Sisto, Santos & Noronha, 2004) para se avaliarem as condutas inteligentes necessárias para respondê-lo. Este estudo, além de fornecer indícios de que uma pontuação única é insuficiente para explicitar os raciocínios exigidos para responder ao R1-Forma B, apresentou dados que alicerçam a necessidade de dois fatores, que seriam suficientes para explicitar os mecanismos psicológicos subjacentes para responder ao teste.

Para os dois fatores propostos no Manual (Sisto, Santos & Noronha, 2004) foram encontrados indícios suficientes para sugerir a unidimensionalidade em ambas as situações. Assim, a escala formada para cada um dos fatores separadamente, com base nos resultados deste estudo, poderia ser interpretada no sentido de que o modelo Rasch retirou de formas bastante similares as variâncias de cada item.

Em razão disso, pode-se afirmar que no, caso dos dois fatores, as respostas das pessoas para todos os itens do teste dependeriam de um único traço subjacente ou habilidade. Dessa forma, a interpretabilidade do fator estaria assegurada, pois uma única habilidade estaria presente quando as pessoas respondessem ao Fator 1, da mesma forma que apenas uma única habilidade, mas diferente da anterior, estaria predominando quando as pessoas estivessem resolvendo os itens do Fator 2. Em consequência, o resultado geral obtido no teste pode ser interpretado, ao mesmo tempo em que se pode conhecer quais mecanismos foram mais usados para as pessoas obterem a pontuação que obtiveram.

REFERÊNCIAS

- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Brooks Cole.
- Almeida, L. S. (1988). *Teorias da inteligência*. Porto: Edições Jornal de Psicologia.
- Alves, I. C. B. (2002). *R1: teste não-verbal de inteligência: manual*. São Paulo: Vetor.
- American Psychological Association - APA (1995). *Standards for educational and psychological tests*. Washington: APA.
- Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society*, 42, 293–321.
- Boccalandro, E. R. (2003). *G36: teste não verbal de inteligência: manual*. São Paulo: Vetor.
- Brody, N. (1992). *Intelligence*. San Diego: Academic Press.
- 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.
- Cunha, S. E. (2001). *Matrizes Progressivas: escala geral*. Rio de Janeiro: CEPA.
- 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: 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: 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*. Mackintosh: Praeger.
- Jensen, A. R. (1994). Charles Edward Spearman (1863-1945). In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 1007-1014). New York: MacMillan.
- Linacre, J. M. (1996). Structure in Rasch residuals: Why principal components analysis? *Rasch Measurement Transactions*, 10(3). Disponível em: <<http://www.rasch.org/rmt/rmt122m.htm>>. (Acesso em 01/12/2003).
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum.
- Lord, F. M. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13, 517–548.
- McDonald, R. P. & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 27, 82–99.
- Noronha, A. P. P. (2001). Análise de coeficientes de testes de inteligência. *Psico*, 32, 73-86.
- Oliveira, R. (1973). *R1: teste não verbal de inteligência - Forma B*. São Paulo: Vetor.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielson & Lydiche.

- Raven, J., Raven, J. C. & Court, J. H. (1991). *Manual for Raven's Matrices e Vocabulary Scales: Section 1. General overview*. Oxford: Oxford Psychologists.
- Ribeiro, I. S. (1998). *Mudanças no desempenho e na estrutura das aptidões*. Tese de Doutorado Não-Publicada. Universidade do Minho. Braga, Portugal.
- 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.
- Sisto, F. F., Codenotti, N., Costa, C. A. J. & Nascimento, T. C. N. (1979). Testes Psicológicos no Brasil: que medem realmente? *Educação e Sociedade*, 1(2), 152-165.
- Sisto, F. F., Santos, A. A. A. & Noronha, A. P. P. (2004). *Teste R-1 de Inteligência Não Verbal – Forma B. Manual*. São Paulo: Vetor.
- Smith, R. M. & Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective Measurement: Theory into Practice* (v. 2). Norwood: 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.
- Sternberg, R. J. & Powell, J. S. (1983). The development of intelligence. In P. H. Mussen (Ed.), *Handbook of child psychology* (pp. 341-419). New York: John Wiley & Sons.
- Sternberg, R. J. (1986). *Intelligence applied: Understanding and increasing your intellectual skills*. New York: Harcourt Brace Javanovich.
- Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38, 406-427.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago.
- Thurstone, L. L. & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago.
- 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, 753-755. Disponível em: <<http://www.rasch.org/rmt/rmt142m.htm>>. (Acesso em 01/12/2003).

Recebido em 30/11/2004

Aceito em 11/07/2006

Endereço para correspondência: Fermino Fernandes Sisto. R. Carlos Guimarães, 150, ap. 82, CEP 13 024-200, Campinas-SP.
E-mail: fermino.sisto@gmail.com