

Um estudo sobre a dimensionalidade do teste do Desenho da Figura Humana

Fermino Fernandes Sisto

Universidade São Francisco, Atibaia-SP

RESUMO

Nesta pesquisa foi avaliado o ajuste do modelo Rasch quanto à unidimensionalidade do teste Desenho da Figura Humana. Foram estudadas 1.207 escolares de ambos os sexos, com idades de 5 a 10 anos. 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. Esses itens foram novamente analisados pelo modelo de Rasch, separadamente, e correlacionadas as pontuações do fator positivo com o negativo. O resultado sugeriu a existência de mais de uma dimensão, em contraposição ao uso de uma pontuação total para a avaliação. O estudo sobre as perturbações de medida relacionadas ao examinando indicou que o principal problema está relacionado à presença ou ausência de detalhes no desenho incompatíveis com as características globais da figura.

Palavras-chave: modelo Rasch; unidimensionalidade; teste do Desenho da Figura Humana.

ABSTRACT

Study about the dimensionality of the Human Figure Drawing test

This work assessed the adjustment of Rasch model concerning the one-dimensionality of the Human Figure Drawing Test. The participants were 1.207 students aged from 5 to 10 years old, of both genders. Considering that residuals may be enough to violate the items intrinsic one-dimensionality, the analysis of residuals principal component was used to assess the possible violation. Items with higher factorial load related to the positive and negative dimension of the first factor were selected. These items were re-analyzed through Rasch model within each group of data, and then the positive and negative factors were correlated. The result suggested the existence of more than one dimension, opposing to the use of a total score for diagnosis. The study about the measure disturbances at the examinees level showed that the main problem is related to the presence or absence of drawing details that are incompatibles to the figure's global characteristics.

Keywords: Rasch model; unidimensionality; Human Figure Drawing test.

Uma das medidas usadas frequentemente na avaliação da inteligência é o Teste do Desenho da Figura Humana (DFH), que é levado em consideração quando se quer avaliar habilidades cognitivas não verbais (Fabry & Bertinetti, 1990). As pontuações nesse instrumento tendem a se correlacionar positivamente com outras medidas de inteligência (correlações que variam do 0,05 ao 0,92, de acordo com Salvia & Ysseldyke, 1985) e fornecem uma estimativa do desenvolvimento intelectual pouco influenciada por diferenças da cultura e da língua.

O Desenho da Figura Humana (DFH) é bastante atrativo para crianças e seu uso constitui uma forma nem ameaçadora, nem invasiva, para avaliar habilidades cognitivas. A tarefa de desenhar uma figura humana é familiar aos membros de todas as culturas e tem-se mostrado independente de experiência acadêmica. A reprodução da figura humana, como um indicador do nível intelectual das crianças, inclui as vantagens de ser aparentemente não viesada e eficiente. É de fácil aplicação porque sua natureza não-verbal e a rapidez com que pode ser realizada a tornam apropriada

para o uso com pessoas com problemas de atenção ou comunicação oral, como também com pessoas com outros tipos de limites (Sisto, 2000).

Apesar de uso difundido, entretanto, os estudos não puderam provar que o DFH mede a inteligência da mesma maneira que as Matrizes Progressivas de Raven, a escala da inteligência de Stanford-Binet, o Porteus Maze Test, o WISC-R, e a Escala de Inteligência de Wechsler para crianças, entre outros, parece capaz de fazer (Harris, 1963; White, 1979; Oakland & Dowling, 1983; Abell, von Briesen & Watz, 1996). Harris (1963) reviu 100 estudos que usaram o DFH, e com base nessa análise concluiu que embora fosse uma medida de confiança para o desempenho das crianças não é um preditor válido de medidas de inteligência; ao lado disso, há pouca evidência empírica que suporta seu uso para avaliar a inteligência.

Além desses estudos, outras investigações relataram resultados que fornecem pouco incentivo para a prática de estimar a inteligência de uma criança pelo DFH (Kamphaus & Pleiss, 1991; Aikman, Belter &

Finch, 1992, entre outros). Em acréscimo, alguns pesquisadores (Motta, Little & Tobin, 1993; Gresham, 1993, por exemplo) discutiram que apesar da validade de seu uso, os resultados mais consistentes obtidos por outras medidas tornaram seu uso supérfluo. Em resumo, as queixas sobre a falta de evidências validade do DFH para avaliações psicológicas são encontradas freqüentemente, mas também foi afirmada muitas vezes.

Essa variabilidade dos resultados das pesquisas que envolvem o teste DFH pode estar relacionada ao que se chama de perturbações de medidas. Essas perturbações normalmente são estudadas em três âmbitos, quais sejam, no âmbito do item, do examinando e da aplicação (Meijer, 1996; Wright & Stone, 1979).

Neste estudo a preocupação se centrará nos dois primeiros âmbitos. Com relação ao segundo, as perturbações de medida relacionadas ao examinando, das oito possibilidades analisadas (Meijer, 1996; Wright & Stone, 1979) serão usadas as propostas por Linacre e Wright (1994), pois eles explicitam critérios estatísticos para elas, como também para os casos considerados modelados pela análise de Rasch. Os critérios de desajuste do protocolo do examinando ao modelo terão como finalidade oferecer as características de desajuste mais comuns e discuti-las.

Por sua vez, as perturbações da medida no âmbito do item englobam preocupações relacionadas à multidimensionalidade, ao viés, a opções de respostas corretas múltiplas e a item com palavras de compreensão ambígua. Esta pesquisa terá como foco a questão relacionada à multidimensionalidade do item, mais especificamente à possibilidade do DFH ser unidimensional.

O estudo psicométrico da inteligência geralmente envolve uma análise da habilidade crescente das crianças para resolver os problemas que a requerem. Unidimensionalidade pode ser compreendida como as respostas a todos os itens do teste sendo dependentes do mesmo traço ou habilidade subjacente. Spearman (1904) foi o primeiro a construir um método estatístico para estudar os testes de inteligência e sua característica principal era buscar o que havia de comum em todos os elementos analisados para sua extração. Em outros termos, construiu um modelo unidimensional de intercorrelações que usou primeiramente para estudar os testes de inteligência.

O presente estudo foi realizado em uma tentativa de verificar se o teste do Desenho da Figura Humana (DFH) pode ser considerado unidimensional. No entanto, neste estudo não se usou o modelo de análise fatorial de Spearman, mas a Teoria de Resposta ao Item, mais especificamente o modelo de Rasch, cujo pressuposto fundamental é a unidimensionalidade, com vistas a estudar o ajustamento dos dados ao modelo.

Um teste psicológico é considerado unidimensional quando seus dados permitem inferir a presença de apenas uma dimensão para explicá-los. Em outros termos, apenas uma única característica ou habilidade de quem realiza o teste estaria envolvida nas respostas. Assim, a unidimensionalidade implica em uma alta consistência interna, pois todos os itens estariam selecionados para medir uma única habilidade. No entanto, não se pode confundir consistência interna com dimensionalidade, porque nem sempre uma alta consistência interna implica em unidimensionalidade.

Tecnicamente falando, um conjunto de itens de um teste é considerado unidimensional quando as distribuições dos itens são independentes e explicáveis por um único traço latente. Por isso, o problema de se saber se um teste é unidimensional tem consequências importantes para sua validade de construto e sua pontuação. Pode-se supor, por exemplo, que o DFH contém dois grupos de itens, nos quais as respostas no primeiro grupo dependem principalmente da “habilidade para desenhar”, enquanto que as respostas no segundo dependem essencialmente de “adição de detalhes evolutivos”. Conseqüentemente, as pessoas com um mesmo resultado podem ter elevada habilidade para desenhar, mas baixa em adições do detalhe ou inversamente, e a possibilidade de combinação dos escores nos dois grupos é muito variada. Dessa forma, para inferir a interpretação teórica adequada da pontuação total será impossível, pois ela significaria uma combinação não identificada de ambas as habilidades. Esse fato enfraquece a validade de construto do teste, como definida pelos Standards (APA, 1985); no entanto, poderia explicar a diversidade de coeficientes de correlações com outros testes da inteligência relatada na literatura. Além disso, na aplicação prática do teste o comportamento de cada pessoa seria sumariado mais apuradamente com duas pontuações diferentes em vez de uma.

A importância de estudos sobre a unidimensionalidade tem um papel importante, pois se relaciona à evidência de validade de constructo do teste, o que foi muito bem apontado por Schmidt e Hunter (1998) em sua revisão de 85 anos de estudos de validade. Uma evidência de validade dessa natureza significa que uma interpretação teórica adequada pode ser dada às pontuações do teste (APA, 1985).

Muitos estudos de validade de constructo consideram que as pontuações do subteste ou as dos itens são relacionadas entre si e, assim, podem ser chamados de estudos da estrutura interna (Cronbach & Meehl, 1955). Como a unidimensionalidade considera as relações entre os itens dentro de um teste, sua avaliação é essencialmente um o estudo da estrutura interna porque é testado se todos os itens medem o mesmo constructo. Assim, a unidimensionalidade significaria

que as respostas das pessoas para todos itens dos testes dependeriam do traço ou da habilidade subjacente.

Muitos métodos podem ser usados para avaliar a dimensionalidade. Hattie (1985) relaciona 85 diferentes formas de se avaliá-la, muitas delas conhecidas como ineficazes. Por um lado, a teoria clássica usa modelos estatísticos nos quais as pontuações dos itens são relacionadas entre si por meio da análise fatorial, com vista a inferir a evidência de validade de constructo (Endler, Parker & Summerfeldt, 1998; entre outros).

No caso do Desenho da Figura Humana, por se tratar de variáveis binárias, os métodos da teoria clássica podem extrair muitos fatores, alguns dos quais são relacionados às dificuldades do item (Hattie, 1985; McDonald & Ahlawat, 1974; por exemplo) e outros fatores artificiais. É também necessário comentar que os modelos da análise fatorial desenvolvidos para tratar esse tipo de variável não deram uma solução satisfatória para o problema do número dos fatores, em razão da suposição dos testes estatísticos usados de que o traço latente subjacente está distribuído normalmente, o que pode não ser válido (Christoffersson, 1975; Bartholomew, 1980).

Por outro lado, existe a alternativa fornecida pelos modelos de Teoria de Resposta ao Item (TRI), desenvolvidos especificamente para variáveis binárias, os quais não necessitam da suposição de que o traço latente esteja distribuído normalmente (Allen & Yen, 1979; Lord, 1953; Lord, 1980; van der Linden & Hambleton, 1997). Desses modelos, para este estudo foi escolhido o de Rasch (1960) por razões técnicas. Uma delas deve-se ao fato de que sua teoria estatística é bem desenvolvida e fácil, comparada com outros modelos de TRI (Fischer, 1995; Glas, 1988; Glas & Verhelst, 1995). Em contraposição, a escolha do modelo de dois parâmetros (uma alternativa) para avaliar os dados fornecidos pelo teste do Desenho da Figura Humana não será feita porque carece de uma rigorosa sustentação matemática que algumas estatísticas do modelo de Rasch têm. Adicionalmente, exatamente onde as violações da unidimensionalidade são mais prováveis, ou seja, nas frequências bivariadas, o modelo logístico de dois parâmetros não possui estatísticas bem desenvolvidas (van den Wollenberg, 1982).

A esses aspectos deve-se somar o fato de que o modelo de Rasch possui um critério muito claro de unidimensionalidade para verificar se os dados se ajustam ao modelo, dificultando a artificialidade. Esse critério é o princípio que norteia a análise dos dados quando se verifica seu ajuste ao modelo, qual seja, é assumido que na resposta de uma pessoa a um item do teste existe alguma quantidade de habilidade presente. Em consequência, se cada pessoa tem alguma quantidade dessa habilidade, ela poderia ser representada

por um valor numérico ou por uma pontuação. E essa pontuação poderia ser colocada em uma escala dessa habilidade. Nesse contexto, é fácil assumir que em cada nível da habilidade existe uma probabilidade, possível de ser calculada, de a pessoa pertencente a esse nível dar uma resposta correta ao item. E a probabilidade seria baixa para a pessoa com pouca habilidade e alta para a pessoa com elevada habilidade.

No modelo de Rasch (Rasch, 1960), a probabilidade da resposta correta aumenta em função da habilidade da pessoa (monotônica na habilidade) assim como ela diminui em função da dificuldade do item (monotônica na dificuldade). Ao mesmo tempo, supõe-se que sua fórmula funcione para todos os sujeitos e todos os itens. Em decorrência, a habilidade de cada pessoa permanece constante em todo o teste ou, em outras palavras, a pessoa usa a mesma habilidade para responder a todos os itens (habilidade unidimensional), e a dificuldade relativa de cada item permanece também a mesma para todas as pessoas estudadas (dificuldade unidimensional).

Com base nesse paradigma, espera-se que os itens possam ser ordenados por suas dificuldades e essa ordem é independente da habilidade da pessoa, isto é, uma pessoa interessada, ou esperta, ou cansada, ou desmotivada não mudaria essa ordem. Ao mesmo tempo, outra decorrência importante, para o processo de avaliação, é a suficiência da pontuação total para o traço subjacente, pois ela seria uma “estatística suficiente” para a habilidade verdadeira subjacente, o que significa que conteria toda a informação sobre a habilidade da pessoa no traço latente em questão. Assim, estaria assegurada que nenhuma informação relevante seria perdida ao se informar somente a pontuação total da pessoa, pois ela não engloba outra informação além da habilidade avaliada, nem informações relevantes estariam deixando de ser informadas.

Se o teste do DFH se ajustar no modelo de Rasch, a pontuação total seria o melhor resumo do padrão da resposta da pessoa, tornando possível inferir a interpretação teórica adequada dessa pontuação. No caso que o teste do DFH não se ajustar ao modelo de Rasch, um único escore não seria uma estatística suficiente para mostrar a quantidade da habilidade subjacente. Isto implicaria na necessidade de mudar o sistema de correção do teste DFH para não negligenciar as dimensões relevantes que são medidas por ele, ou encontrar uma forma que se ajustasse à unidimensionalidade.

Com base nas ponderações feitas, é necessário deixar claro que este estudo não tem o intuito de comprovar qualquer teoria psicológica, mas foi projetado por considerações teóricas da medida, no sentido de determinar quão válida é a interpretação sobre a pontuação da pessoa. E valendo-se dessas ponderações, este estudo avalia o ajuste do modelo de Rasch em relação à unidimensionalidade do DFH.

De certa maneira, a suposição de unidimensionalidade do modelo é analisada focalizando na suficiência da pontuação total e no aumento monotônico e paralelo da função da resposta ao item. Para verificar esse ajuste, será usada a análise de componentes principais dos resíduos (Linacre, 1996), em detrimento de vários outros testes estatísticos (van der Wollenberg, 1982; Glass, 1988, 1989; por exemplo). Essa análise supõe que o modelo de Rasch extrai dos dados sua melhor dimensão linear, mas supõe também a possibilidade que cada observação abrange ainda, em algum nível, informação específica dessa observação que poderia contradizer a unidimensionalidade intrínseca do item. Uma maneira de conhecer essas informações específicas seria analisando as diferenças entre o que o modelo de Rasch prevê e o dado observado, ou seja, os resíduos. O que a análise de componentes principais dos resíduos detectará são as características comuns entre os itens, procurando por indicações de estruturas secundárias ou de subdimensões nos dados.

MÉTODO

Participantes

Foram pesquisadas 1.207 crianças (50,54% ou 610 meninos e 49,46% ou 597 meninas), freqüentando desde a pré-escola até a 4ª série do ensino fundamental, com idades de 5 a 10 anos (média= 8,14 e desvio padrão de 1,42), oriundas de vários estratos socioeconômicos. Três critérios foram usados para excluir protocolos, conforme a indicação de Goodenough (1928), quais sejam, ausência de cabeça, perna ou braço. Assim, foram eliminados 7 (0,58%) protocolos por ausência de cabeça, 16 (1,33%) protocolos com apenas cabeça, 17 (1,41%) protocolos com apenas cabeça mais um dos outros dois critérios, e 149 (12,34%) protocolos por não atenderem aos três critérios. Assim, das 1.396 crianças que fizeram os desenhos foram eliminados 189 protocolos, restando 1.207 para este estudo. A Tabela 1 apresenta a distribuição dos participantes por idade, série e sexo.

Tabela 1: Freqüência de participantes por idade, série e sexo

Séries	maternal		Pré-escola		1ª. série		2ª. série		3ª. série		4ª. série		Total
Sexo	M	F	M	F	M	F	M	F	M	F	M	F	
Idades	5	29	22	-	-	-	-	-	-	-	-	-	51
	6	10	8	28	24	5	5	-	-	-	-	-	80
	7	-	-	7	7	141	126	7	6	-	-	-	294
	8	-	-	-	-	47	31	110	115	3	7	-	313
	9	-	-	-	-	1	1	36	24	95	125	6	298
	10	-	-	-	-	-	-	9	4	33	38	43	171
Total	39	30	35	31	194	163	162	149	131	170	49	54	1207
		69		66		357		311		301		103	

Teste da Figura Humana

A instrução para aplicação do Teste do Desenho da Figura Humana (DFH) foi para que a criança desenhasse a figura de uma pessoa com a maior quantidade de detalhes que pudesse, usando um lápis e uma folha de papel sulfite. Foi permitido o uso de borracha na confecção do desenho. A aplicação foi coletiva a todas as crianças de cada sala de aula.

Com base em 51 dos detalhes ou critérios de Goodenough (1928) os protocolos foram corrigidos. Os detalhes foram os seguintes: (1) cabeça, (2) pernas, (3) braços, (4) tronco, (5) tronco comprido, (6) ombros, (7) união braços pernas, (8) união braço/perna na posição correta, (9) pescoço, (10) pescoço em continuação da cabeça, (11) olhos, (12) nariz, (13) boca, (14) boca/nariz em duas dimensões, (15) orifícios do nariz, (16) cabelos, (17) cabelos além do crânio, (18) roupa, (19) pelo menos duas peças de roupa, (20) roupa sem transparência, (21) quatro peças de roupa bem definidas, (22) roupa completa sem transparência, (23) dedos, (24) número correto de dedos, (25) detalhes de dedos, (26) polegar em oposição, (27) representação

mão-dedo/braço, (28) articulação braço-perna-cotovelo, (29) articulação perna-jelho-cadeira, (30) proporção cabeça, (31) proporção braços, (32) proporção pernas, (33) proporção pés, (34) proporção braço-pernas em duas dimensões, (35) representação saltos, (36) coordenação motora linhas A, (37) coordenação motora linhas B, (38) coordenação motora cabeça, (39) coordenação motora tronco, (40) coordenação motora braço-perna, (41) coordenação motora face, (42) orelhas, (43) orelhas posição-proporção, (44) sobancelha/pestana, (45) pupilas, (46) olho proporção, (47) olhar, (48) testa e queixo, (49) queixo, (50) perfil A, (51) perfil B.

Desses critérios, três deles não foram considerados para análise, quais sejam, cabeça, perna e braço, em decorrência de terem sido usados como critérios de exclusão de protocolos, dos que restaram todos apresentavam as três informações. Assim, esses itens não forneciam variabilidade para qualquer análise. Dois dos detalhes ou critérios de correção (perfil) não registraram ocorrência.

Em resumo, de fato, foram usados 46 itens para estudo e a variabilidade possível do instrumento seria de 0-46. Os itens foram pontuados com o valor de um quando o detalhe estava presente e com zero quando o detalhe estava ausente.

Procedimentos de análise de dados

Primeiramente foram computados média, desvio padrão, valor mínimo e máximo e consistência interna (alfa de Cronbach) para caracterização dos participantes em estudo. Depois foram eliminados os itens acertados por todos os participantes, cabeça, perna e braço e aqueles que tiveram uma quantidade de acerto bastante baixa, qual seja, os referentes ao perfil.

Em seguida, 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 forneceram um eigenvalue superior a 1,4 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 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.

Além do mais, os padrões de ajuste das crianças serão classificados de acordo com os *infit* e *outfit* (Linacre & Wright, 1994). É necessário salientar que o ideal seria que os valores de ajuste do *infit* e *outfit* fossem iguais a 1. Entretanto, isso é bastante complicado, principalmente quando se trata de grande quantidade de dados. Assim, propuseram uma tolerância de 20% para mais e para menos desse valor, fornecendo um intervalo de ajuste.

Em uma adaptação da proposta desses autores para o caso específico do teste DFH, considerou-se que a criança estava ajustada ao padrão quando tanto o *infit* quanto o *outfit* tiveram um valor entre 0,80-1,2. Para as situações de desajuste ao modelo, classificou-se como descuido, quanto o valor do *infit* estava ajustado (0,80-1,2) e o *outfit* acima do intervalo considerado ajustado, ou seja, valores superiores a 1,2; como desajuste ao crivo de correção, quando tanto o *infit* quanto o *outfit* tiveram um valor acima do intervalo

considerado ajustado (0,80-1,2); como sugerindo conhecimento específico, se o *infit* foi superior ao intervalo considerado ajustado, mas o *outfit* dentro do parâmetro de ajuste; e, por fim, determinista, quando a tendência se enquadrou no padrão Guttman, ou seja, tanto o *infit* quanto o *outfit* tiveram um valor inferior ao intervalo de ajuste.

RESULTADOS

Tabela 2: Estatísticas descritivas do teste do Desenho da Figura Humana

Idades	Alfa de Cronbach	Valor mínimo	Valor máximo	Média	Erro padrão	Desvio padrão
todas	0,88	3	49	20,45	0,1400	7,16
5	0,84	3	27	11,00	0,4753	4,68
6	0,79	5	27	13,54	0,3565	4,4
7	0,82	3	38	17,89	0,2287	5,74
8	0,84	3	41	20,74	0,2334	6,09
9	0,84	7	49	22,53	0,2490	6,31
10	0,89	6	45	25,35	0,3775	7,74

Também, os coeficientes de consistência interna (alfa) variaram entre 0,79-0,89, sendo que apenas um foi inferior a 0,80. Assim, pode-se considerar que a precisão do instrumento está em um nível muito bom.

O resultado da análise por componentes principais dos resíduos indicou que o fator 1 explicou 2,5 de 48

unidades de variância residual, um por item. A proporção do fator no Yardstick foi de 263,6/1. Optou-se por continuar a análise, pois o eigenvalue foi superior a 1,4. Os resultados do fator 1 estão na Tabela 3. Foram fornecidos apenas os dados dos itens que obtiveram uma saturação superior a 0,20.

Tabela 3: Estatísticas do fator 1, extraído por análise de componentes principais dos resíduos, com saturação acima de 0,20

saturação fatorial	measure	infit MNSQ	outfit MNSQ	item
0,53	1,47	0,97	0,77	roupa sem transparência
0,51	-0,23	0,86	0,79	roupa pelo menos 2 peças
0,41	2,81	0,99	0,67	roupa 4 peças bem definidas
0,39	-0,77	1,08	1,09	coordenação motora linhas A
0,38	-1,89	0,85	0,75	roupa
0,22	0,68	1,02	0,97	ombros
0,20	1,38	1,25	1,49	articulação perna Joelho-cadeira
-0,47	2,30	0,91	0,61	dedos polegar em oposição
-0,47	2,08	0,93	0,72	dedos detalhes
-0,34	0,55	1,11	1,31	dedos representação mão-dedo/braço
-0,34	0,61	0,98	0,92	proporção braços
-0,30	1,91	0,90	0,73	coordenação motora faces
-0,30	0,93	0,92	0,84	número correto de dedos
-0,29	1,99	0,87	0,76	coordenação motora cabeça
-0,28	1,48	0,76	0,49	coordenação motora braço-perna
-0,22	-2,25	1,00	1,01	dedos

Os itens com saturação fatorial positiva escolhidos foram roupa sem transparência, roupa com pelo menos duas peças, roupa com quatro peças bem definidas e coordenação motora linhas A. Os com saturação fatorial negativa escolhidos foram dedos com polegar em oposição, dedos detalhes, dedos representação

mão-dedo/braço e proporção dos braços. Esses itens positivos e negativos foram submetidos, separadamente, a uma nova análise pelo modelo do Rasch e correlacionados os measures dos itens positivos e negativos de cada pessoa.

O coeficiente correlação de Pearson foi $r=0,125$, que corrigido pela fórmula de profecia de Spearman-Brown resultou em um valor $r=0,44$. Esse valor é suficientemente baixo para indicar que os dados possuem mais de uma dimensão. Aparentemente, os detalhes relacionados com o corpo são comparados com as roupas, sugerindo dois tipos de habilidades.

A análise seguinte trata dos ajustes estatísticos das crianças em relação à escala construída pelo modelo. Para essa análise são levados em consideração os valores do *infit* e do *outfit*. Deve-se considerar que o valor do *outfit* sofre a influência de respostas inesperadas,

remotas, fora de propósito, sendo sensível a elementos descontextualizados. Por sua vez, o valor do *infit* é dominado por padrões internos e inesperados da pessoa, tais como informações objetivas ou conhecimento adquirido, sendo sensível a elementos apreendidos anteriormente.

Com base nos dados de *infit* e *outfit*, as respostas das crianças foram analisadas e foram encontrados 220 (18,23%) protocolos que não se ajustaram à escala extraída pelo modelo Rasch. A classificação dos padrões encontra-se na Tabela 4.

Tabela 4: Frequências e porcentagens dos protocolos em relação aos padrões de ajuste-desajuste ao modelo

Padrões	<i>Infit</i>	<i>Outfit</i>	Frequência	Percentual
1 – ajustado ao modelo	0,80-1,2	0,80-1,2	987	81,77
2 – descuido	0,80-1,2	> 1,2	10	0,83
3 – desajuste ao crivo de correção	> 1,2	> 1,2	119	9,86
4 – conhecimento específico	> 1,2	0,80-1,2	8	0,66
5 – determinista (Guttman)	< 0,80	< 0,80	83	6,88
Total			220	18,23

CONSIDERAÇÕES FINAIS

A indicação para o diagnóstico dos examinandos por meio do teste do Desenho da Figura Humana sempre foi o de que eles deveriam ser avaliados por uma única pontuação (a soma de todos os itens corretos), explicitando a pressuposição de que se trataria de uma medida unidimensional. No entanto, não pôde ser comprovada a unidimensionalidade do teste DFH analisada pelo modelo de Rasch no teste como um todo.

Os indícios sugerem a presença de mais de uma dimensão, o que, de forma prática, implicaria na necessidade de se determinar os fatores presentes, sua interpretabilidade e de se ter mais de uma pontuação para descrever as condutas ou habilidades cognitivas que as pessoas lançam mão quando estão realizando o desenho de uma pessoa. Em um certo sentido, é possível interpretar esse resultado como uma possibilidade de explicação para as conclusões de Harris (1963), White (1979) Oakland e Dowling (1983) e Abell, von Briesen e Watz (1996), entre outros, quando afirmaram que o DFH não mediria a inteligência da mesma maneira que Matrizes Progressivas de Raven, a escala da inteligência de Stanford-Binet, o Porteus Maze Test, o WISC-R, e a Escala de Inteligência de Wechsler para crianças. Nesse contexto, pesquisas seriam necessárias para se avaliar o quanto poderia ser medido dessas provas pela DFH e o que mais ele estaria avaliando, além desse tipo de habilidade cognitiva.

De fato, é fácil concordar com Kamphaus e Pleiss (1991) e Aikman, Belter e Finch (1992) que as pes-

quisas em busca de um instrumento paralelo ou muito próximo a isso forneceram pouco estímulo para o uso do DFH em termos de diagnósticos. No entanto, dada as características do DFH seria importante encontrar evidências de validade para ele, pois dificilmente se disporá de um instrumento não invasivo, rápido, não ameaçador, sem necessidade de adaptação para pessoas de diferentes culturas, níveis socioeconômicos e acadêmicos, sem necessidade de domínio da língua ou outros problemas que não envolvam questões psicomotoras. Ao lado disso, o fato de apenas em 0,8% das crianças ter sido encontrado um efeito que pode ser atribuído a uma habilidade ou conhecimento específico, lança muita dúvida sobre a possibilidade de uma habilidade para desenhar em certas crianças facilitar ou melhor pontuação no DFH.

E uma forma de se fazer isso talvez fosse mudando a perspectiva na qual tem sido avaliado, ou seja, concebendo-o como multidimensional. As análises sugeriram que a unidimensionalidade do DFH carece de sustentação e a pontuação total, como foi proposta pelo autor do instrumento para sua avaliação, pode não ser uma medida suficiente das habilidades envolvidas para as pessoas responderem ao instrumento.

Acrescente-se a isso que é necessário considerar que talvez não haja necessidade de se utilizar todos os 51 indicadores propostos por Goodenough (1928), ou os 46 usados neste estudo. Na realidade, a re-análise feita por Harris (1963) aumentou a quantidade de itens e essa escala tem sido usada na maior parte das pesquisas citadas neste estudo. E a questão pode não ser a

necessidade de aumentar os indicadores, mas diminuí-los, para que meçam muito mais o desenvolvimento intelectual ou cognitivo do que parece estar medindo.

A esse respeito, os dados relativos aos protocolos que não poderiam ser avaliados pela escala criada pelo modelo de Rasch podem fornecer alguma informação a esse respeito. Do total de protocolos estudados, aproximadamente 82% se ajustaram ao modelo. Quase 7% se enquadraram na categoria determinista, pois segue o padrão de Guttman, o que não causa qualquer problema de interpretabilidade dos protocolos: eles apenas diminuem a precisão da medida. Por isso, tampouco precisam ser considerados problemáticos em relação à validade. Restam, assim, quase 11% de protocolos que estariam causando perturbações na medida.

A categoria que merece atenção no caso desses dados é a três, denominada de desajuste ao crivo de correção, responsável por 9,86% dos protocolos que fogem ao padrão. A característica básica desses protocolos foi apresentar detalhes não esperados em relação à dificuldade dos itens, ou seja, ou foram desenhados detalhes raramente colocados em desenhos do padrão em questão, ou não foram delineados detalhes comuns a desenhos no padrão do realizado. Esse dado parece sugerir que nem todos os detalhes usados para avaliar o DFH são evolutivos, a despeito dos principais estudos sobre evidências de validade usarem a diferenciação por idades como critério.

Neste estudo, as análises sugeriram que a unidimensionalidade do teste DFH 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. Por ser um instrumento proposto para avaliação da inteligência de crianças ou de pessoas com baixo nível de escolaridade ou com pouco conhecimento da língua portuguesa é necessária muita cautela para seu uso diagnóstico, tal como os critérios estão postos. Para futuras revisões do instrumento, pode ser desejável criar diferentes escalas com interpretações do tipo de inteligência envolvida ou dos mecanismos psicológicos subjacentes. Ao lado disso, podem ser necessários também estudos sobre os itens, pois todos eles podem não ser necessários para uma avaliação de habilidades cognitivas das pessoas.

REFERÊNCIAS

- Abell, S. C., Von Briesen, P. D. & Watz, L. S. (1996). Intellectual evaluations of children using human figure drawings: an empirical investigation of two methods. *Journal of Clinical Psychology*, 52(1), 67-74.
- Aikman, K. G., Belter, R. W. & Finch, A. J. (1992). Human figure drawings: validity in assessing intellectual level and academic achievement. *Journal of Clinical Psychology*, 48(1), 114-120.
- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Brooks/Cole, Monterey, CA.
- APA (1985). American Psychological Association, American Educational Research Association & National Council on Measurement in Education. *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.
- Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40, 5-32.
- 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.
- Fabry, J. J. & Bertinetti, J. F. (1990). A construct validation study of the Human Figure Drawing Test. *Perceptual and Motor Skills*, 70(2), 465-466.
- Fischer, G. H. (1995). Derivations of the Rasch model. Em G. H. Fischer & I. W. Molenaar (Orgs.), *Rasch models. Foundations, recent developments, and applications* (pp. 15-38). 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.
- Glas, C. A. W. & Verhelst, N. D. (1995). Testing the Rasch model. Em G. H. Fischer & I. W. Molenaar (Orgs.), *Rasch models. Foundations, recent developments, and applications* (pp. 69-96). New York: Verlag.
- Goodenough, F. L. (1928). *Measurement of intelligence by drawings*. New York: Harcourt, Brace & World.
- Gresham, F. M. (1993). "What's wrong in this picture?": Response to Motta et al.'s review of human figure drawings. *School Psychology Quarterly*, 8(3) 182-186.
- Harris, D. B. (1963) *Children's drawings as measure of intellectual maturity*. New York: Harcourt, Brace, Jovanovich.
- Hattie, J. A. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.
- Kamphaus, R. W. & Pleiss, K. L. (1991). Draw-A-Person techniques: Tests in search of a construct. *Journal of School Psychology*, 29(4) 395-401.
- Linacre J. M. & Wright B. D. (1994). Chi-Square Fit Statistics. *Rasch Measurement Transactions* 8:2 p. 350. Obtido em 10 de dezembro de 2003 do World Wide Web: <http://209.238.26.90/rmt/rmt82a.htm>.
- Linacre, J. M. (1996). Structure in Rasch residuals: Why principal components analysis? *Rasch Measurement Transactions*, 10(3). Obtido em 01 de dezembro de 2003 do World Wide Web: <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, Hillsdale.
- McDonald, R. P. & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 27, 82-99.
- Meijer, R. R. (1996). Person-fit research: an introduction. *Applied Measurement in Education*, 9, 3-8.

- Motta, R. W., Little, S. G. & Tobin, M. I. (1993). A picture is worth less than a thousand words: response to reviewers. *School Psychology Quarterly*, 8(3) 197-199.
- Oakland, T. & Dowling, L. (1983). The Draw-A-Person Test: Validity properties for nonbiased assessment. *Learning Disability Quarterly*, 6(4) 526-534.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielson & Lydiche.
- Salvia, J. & Ysseldyke, J. (1985). *Assessment in special and remedial education*, Boston, MA: Houghton Mifflin.
- 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. (2000). Relationships of the Piagetian cognitive development to Human Figure Drawing. *Journal of School Psychology*, 30(4), 432-441.
- Smith, R. M. & Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. Em M. Wilson (Org.), *Objective Measurement: Theory into Practice* (132-154). Norwood NJ: Ablex.
- Spearman, C. (1904). "General Intelligence", objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- 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. (Orgs.) (1997). *Handbook of modern item response theory*. New York: Springer.
- White, T. H. (1979). Correlations among the WISC-R, PIAT, and DAM. *Psychology in the Schools*, 16(4), 497-501.
- Wright, B. D. & Stone, M. H. (1979). *Best test design: rasch measurement*. Chicago: MESA Press.
- Wright, B. D. (2000). Conventional factor analysis vs. Rasch residual factor analysis. *Rasch Measurement Transactions*, 14(2), p. 753. Obtido em 01 de dezembro de 2003 do World Wide Web: <http://www.rasch.org/rmt/rmt 142m.htm>.

Enviado: 05/03/2005

Revisado: 30/06/2005

Aceito: 10/07/2005

Sobre o autor:

Fermino Fernandes Sisto: Doutor pela Universidad Complutense de Madrid, Livre-docente pela Unicamp e docente do curso de Psicologia e do Programa de Estudos Pós-graduados em Psicologia da Universidade São Francisco, *campus* Itatiba/SP. Endereço eletrônico: fermino.sisto@gmail.com.