

---

# O QUE MEDE O DESENHO DA FIGURA HUMANA? ESTUDOS DE VALIDADE CONVERGENTE E DISCRIMINANTE<sup>1</sup>

---

CARMEN E. FLORES-MENDOZA, MARCELA MANSUR-ALVES

Universidade Federal de Minas Gerais · Brasil

FRANCISCO J. ABAD

Universidad Autónoma de Madrid · Espanha

ÁLVARO JOSÉ LELE

Centro Universitário Newton Paiva · Brasil

## RESUMO

---

*O Desenho da Figura Humana (DFH) constitui um dos instrumentos mais utilizados na prática de avaliação psicológica no Brasil. São poucos os estudos de validade como instrumento de avaliação cognitiva. O presente trabalho analisa o desempenho de crianças belo horizontinas no DFH (sistema Wechsler de correção) e nos testes Matrizes Progressivas de Raven, Escala Verbal do WISC-III, Teste R-2 e Teste de Bender (sistema Koppitz). Os resultados de correlação parcial (controlando-se a idade das amostras) apontaram baixas associações, porém significativas, entre o DFH e os testes Raven (0,204); Escala Verbal do WISC-III (0,299); R-2 (0,330) e Bender (-0,396). Por outro lado, ao comparar três sistemas de correção do DFH (Wechsler, Goodenough e Harris), encontraram-se altos coeficientes de correlação entre eles (0,72 a 0,79). Conclui-se que o DFH é uma medida “proxy” (intermediária) da inteligência e que demanda, em intensidade baixa a moderada, tanto componentes cognitivos quanto psicomotores. Também, observa-se não haver diferença substancial entre os três sistemas de correção, abrindo a possibilidade de usar um ou outro sistema.*

**Palavras-chave:** Desenho da Figura Humana; validade convergente-divergente, Teste R-2; Teste de Raven, WISC-III.

## ABSTRACT

---

**WHAT DOES THE HUMAN FIGURE DRAWING TEST MEASURE? CONVERGENT AND DISCRIMINANT VALIDITY STUDY**  
*The Human Figure Drawing Test (HFD) constitutes an instrument often used in Brazilian psychological assessment practices. Few studies of its validity as a cognitive assessment instrument can be found. The present paper analyses the performance of several samples of children from Belo Horizonte on the HFD (Wechsler scoring system) and on others tests such as Raven’s Coloured Progressive Matrices, verbal scale of WISC-III, R-2 Test, and Bender Test (Koppitz system). The results of partial correlation (controlling age) indicated low, but significant values, between HFD and Raven (.204); IQ verbal (.299); R-2 (.330), and Bender (-.396). In addition, three scoring systems of HFD (Wechsler, Goodenough, and Harris systems) were compared. High coefficients were found (.72 to .79). It is concluded that the HFD is a proxy measure of intelligence and is fairly associated to cognitive and psychomotor abilities. On the other hand, there are not significant differences between the three scoring systems which means that it is possible to use either of them.*

**Key words:** Human Figure Drawing Test; convergent-discriminant validity; R-2 Test, Raven Test; WISC-III.

1. Agradecimento: O presente trabalho recebeu apoio financeiro da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG – Processo nº SHA 80257/03).

## INTRODUÇÃO

A produção gráfica infantil, especialmente no que se refere ao Desenho da Figura Humana, foi sistematicamente estudada no início do século passado nos Estados Unidos a partir do trabalho de Florence Goodenough, o qual deu origem à criação de um instrumento de avaliação psicológica, o chamado *Draw-a-Man Test* ou Teste do Desenho da Figura Humana (DFH) (Alves, 1979). Posteriormente surgiram novas propostas de correção do DFH para utilização em avaliação cognitiva como, por exemplo, o sistema de Harris na década de 1960 (*The Goodenough-Harris Draw-a-Man Test*) e a proposta de Naglieri na década de 1980 (*The Draw a Person: A Quantitative Scoring System*). Desde então, diversos trabalhos foram realizados na tentativa de verificar a relação entre o desempenho gráfico e outros aspectos do desenvolvimento infantil, especialmente aqueles relacionados à inteligência (Kamphaus e Pleiss, 1991).

Abou-Jamra e Castilho (1987) informam sobre antigos estudos de validade do sistema de Goodenough em que se obtiveram correlações entre 0,55 a 0,86 com o Teste Stanford-Binet e de 0,40 com o fator de raciocínio do Teste de Habilidades Mentais Primárias de Thurstone. No caso do sistema Harris, encontraram-se correlações flutuantes. Elas foram desde um nível baixo (0,22 com as Matrizes Progressivas de Raven), até um nível médio (0,40 a 0,54 com o Stanford-Binet) ou alto (0,81 com o WISC).

Kamphaus e Pleiss (1991) citam também uma lista de estudos de validade com resultados variados. Por exemplo, os índices de correlação entre o sistema Goodenough e o Stanford-Binet seriam maiores (variando entre 0,36 a 0,65 dependendo do estudo) do que os encontrados nas investigações que compararam o sistema Harris com outros testes de inteligência (variando entre 0,22 a 0,40). No entanto, deve-se ressaltar que os estudos realizados com o sistema Goodenough foram feitos nas décadas de 1940 e 1950, enquanto que os estudos que utilizaram o sistema Harris foram realizados no final da década de 80, provavelmente com uso de maiores recursos de registro e análise de dados. No que se refere ao sistema de Naglieri, este também apresentou correlações bastante instáveis com testes como o *Kaufman Test of Educational Achievement* (dependendo do estudo, entre 0,0 a 0,47) e o *Matrix Analogies Test-Short-Form* (0,22 a 0,50).

Em geral, uma observação mais cuidadosa dos últimos estudos que empregaram sistemas de correção do DFH permite inferir que as correlações entre essa técnica e outros testes de inteligência são quase sempre menores dos que se encontram entre os testes tradicionais de avaliação cognitiva. Com efeito, no processo de adaptação do WISC-III ao contexto brasileiro (Wechsler, 2002) foi encontrada uma correlação de 0,65 entre o QI Verbal do WISC-III e o Teste das Matrizes Progressivas Coloridas de Raven.

No Brasil, o sistema mais recente de avaliação do Desenho da Figura Humana constitui o trabalho de Wechsler (2003). Trata-se de uma proposta baseada nos trabalhos de Harris, Koppitz e Naglieri. Em um primeiro estudo, a autora apresenta a análise dos desenhos produzidos por crianças oriundas da cidade de Brasília-DF (n=2.391) e da cidade de Campinas-SP (n=588). Em um segundo estudo, estuda o desempenho de 255 crianças de Campinas (SP), Riberão Preto (SP) e Passo Pundo (RS) e, posteriormente, compara o desempenho de 54 crianças brasileiras com o de 55 crianças argentinas. Finalmente, em um terceiro estudo, apresenta uma repadronização do instrumento com uma amostra de 3.340 crianças oriundas de cinco Estados do Brasil.

---

Em todos os estudos, Wechsler (2003) mostra evidências de mudanças desenvolvimentais no instrumento (maior idade, maior pontuação), o que, no seu entender, confirmaria a validade de construto. Com relação a este aspecto, deve-se alegar que, de fato, um instrumento que identifique mudanças de desempenho em função da idade pode estar solicitando traços psicológicos relacionados à inteligência, posto que esta apresenta características inequivocamente desenvolvimentais. Contudo, tal evidência não é suficiente para estabelecer a validade de construto uma vez que outros, relativamente independentes da inteligência, também apresentam mudanças desenvolvimentais (ex. desenvolvimento viso-motor, social, moral, perceptual ou mnêmico).

Com vista a um maior suporte das evidências de validade do instrumento, Wechsler (2003) apresenta os resultados obtidos da comparação entre o desenho da figura humana e o Teste de Integração Viso-Motora de Berry. Encontra uma correlação de 0,59 para o desenho da figura feminina e de 0,62 para a figura masculina. Posteriormente, apresenta um pequeno estudo de validade convergente realizado com 103 crianças, em que correlaciona o DFH com um teste cognitivo baseado nas Matrizes Progressivas de Raven, o chamado Teste Não Verbal de Raciocínio Infantil (TNVRI). Obteve uma correlação de 0,27 para a figura feminina e de 0,21 com a figura masculina. Ainda que se considerem as conclusões da autora de que “*a ampliação dos estudos com o nosso sistema ... confirmaram a relevância do desenho para avaliação cognitiva*” (p. 37), tais resultados sugerem, em realidade, que o instrumento demanda mais aspectos relacionados à coordenação viso-motora do que ao raciocínio.

Os poucos registros internacionais, geralmente antigos, indicam flutuação das associações dos diversos sistemas de pontuação do DFH (embora, entre eles, os coeficientes de correlação sejam altos) e diversas outras medidas cognitivas. Por outro lado, no início de 2000 os estudos sistemáticos realizados por Wechsler no contexto brasileiro utilizaram um teste psicomotor (Teste de Integração Viso-Motora de Berry) pouco conhecido no campo acadêmico e um teste cognitivo (TNVRI) cuja disponibilidade, naquela época, estava restrita ao laboratório que o criou<sup>2</sup>.

Recentemente, Wechsler e Schelini (2006) publicaram um estudo de validade da Bateria Woodcock-Johnson-III. Os resultados indicaram que entre os 10 subtestes, cinco se associaram positivamente (0,358 a 0,512) com o desenho da figura masculina do DFH, quando as crianças eram do sexo feminino, e quatro desses mesmos subtestes tiveram também associação positiva moderada (0,403 a 0,625) com o DFH, quando as crianças eram do sexo masculino. Tais coeficientes parecem mostrar que o DFH pode compartilhar, de fato, aspectos cognitivos demandados pelos testes tradicionais de inteligência. O problema, entretanto, é que o referido estudo não controlou o efeito da idade. Assim, é provável que o tamanho das correlações tenha sofrido influência da variável idade, um fator desenvolvimental compartilhado por ambos os testes. Dessa forma, nos valores das correlações se confundem o efeito da idade e do tipo de processamento mental demandados pelos testes.

Considerando-se que o DFH é uma medida bastante utilizada pelos psicólogos brasileiros, é necessário ampliar as investigações desse instrumento comparando-o a outros tradicionais e fartamente estudados no campo da avaliação cognitiva. Assim sendo, o presente trabalho objetivou levantar informação da natureza do construto subjacente ao Desenho da Figura Humana, especificamente o desenho da figura masculina, por meio de estudos de validade convergente e divergente.

## MÉTODO

### Participantes

Participaram do estudo duas amostras perfazendo um total de 628 crianças. A primeira amostra foi constituída de 335 crianças (171 meninos e 164 meninas) freqüentadoras de uma escola pública de Belo Horizonte, com idade média de 8,9 (DP=1,39; amplitude de idade entre 7 e 11 anos) e a segunda amostra foi constituída de 293 crianças (157 meninos e 136 meninas), com idade média de 8,13 (DP=1,88; amplitude de idade entre 5 e 11 anos) freqüentadoras de três escolas públicas e de três escolas particulares da região de Belo Horizonte.

### Materiais e Procedimento

Ambas as amostras realizaram o desenho da figura masculina seguindo as orientações contidas no manual do teste (Wechsler, 2003). Foi administrado à primeira amostra o Teste das Matrizes Progressivas Coloridas de Raven - Escala Especial (Angelini, Alves, Custódio, Duarte e Duarte, 1999) e a Escala Verbal do WISC-III (Wechsler, 2002) seguindo, também, as instruções contidas nos manuais desses instrumentos. A coleta de dados realizou-se entre final do ano de 2002 e início de 2003. O projeto foi aprovado pelo comitê de ética da universidade e pelo colegiado da escola de ensino fundamental, local de realização do estudo, no primeiro semestre de 2002.

Aplicou-se à segunda amostra o Teste R-2 (Rosa e Alves, 2000) e o Teste Percepto-Visomotor Bender (Koppitz, 1989). A coleta de dados se realizou entre o final de 2004 e início de 2005. As autoridades das escolas de ensino fundamental, pública e particular, autorizaram o processo de coleta de dados. A correção dos desenhos da primeira e segunda amostra esteve a cargo de uma equipe de estudantes do curso de Psicologia previamente treinados pela coordenação do projeto.

Todavia, selecionou-se da primeira amostra uma pequena amostra de 107 desenhos de crianças de três grupos etários e que freqüentavam a mesma escola. Os desenhos foram avaliados por oito alunos do curso de Psicologia, em que dois deles avaliaram sob o sistema Wechsler (DFH), dois sob o sistema Goodenough e quatro sob o sistema Harris.

### Técnicas de análises

O tratamento dos dados utilizou as técnicas de correlação de Pearson e correlação parcial. Também, com o objetivo específico de identificar, se dois coeficientes de correlação do mesmo fenômeno advindos de duas amostras são significativamente diferentes aplicou-se a fórmula de Cohen (1988):

$$Z_{\text{obs}} = Z_1 - Z_2 / \sqrt{[1/N_1 - 3] + [1/N_2 - 3]}$$

Onde  $Z$  é a forma padronizada dos coeficientes de correlação e  $N$  é o tamanho amostral onde foi obtida a correlação. Se o valor  $Z_{\text{obs}}$  observado encontra-se entre  $-1,96$  e  $+1,96$  então as diferenças entre os dois coeficientes de correlação não são estatisticamente significativas. No presente estudo, a fórmula foi empregada para identificar diferenças entre as correlações dos testes DFH e Bender obtidas em vários grupos etários.

---

Para identificar o grau de similaridade entre três variáveis (ou coeficientes de correlação) obtidos em uma mesma amostra utiliza-se a fórmula de Steiger (1980):

$$T = (r_{xy} - r_{xz}) \sqrt{\frac{(n-1)(1+r_{yz})}{2|R| \left[ \frac{n-1}{n-3} + \frac{(r_{xy} + r_{xz})^2}{2} (1-r_{yz})^3 \right]}}$$

No presente estudo, a fórmula foi aplicada para identificar diferenças significativas entre coeficientes de correlação dos três instrumentos aplicados tanto na primeira amostra (DFH/Raven/Escala Verbal WISC-III) quanto na segunda amostra (DFH/Bender/R-2).

## RESULTADOS

O sistema de avaliação Wechsler (2003) para o DFH resgata alguns critérios do sistema Harris e de Naglieri, que por sua vez representa um melhoramento do sistema original proposto por Florence Goodenough. Supõe-se, portanto, haver uma associação, em algum grau, entre os três sistemas de correção. A Tabela 1 apresenta os resultados nessa direção.

**Tabela 1. Correlação entre os sistemas de correção do Desenho da Figura Humana (Sistema Wechsler, Goodenough e Harris).**

Idades	N	DFH/Goodenough	DFH/Harris	Goodenough/Harris
7	39	0,76*	0,52*	0,72*
9	37	0,78*	0,69*	0,75*
11	31	0,69*	0,84*	0,85*
Total	107	0,72*	0,79*	0,72*

Total = controlando-se a idade

\*n.sig<0,001

As altas correlações obtidas nos três grupos etários entre os três sistemas de avaliação do DFH (Wechsler, Goodenough e Harris), sistemas estes empregados por avaliadores diferentes, indicam haver forte similaridade entre eles. As correlações se tornam mais similares quando se controla as idades.

### Primeira amostra: Validade convergente

As estatísticas descritivas (média e desvio padrão) para os testes DFH, Raven - Escala Colorida e Escala Verbal do WISC-III aplicados na primeira amostra (n=335) são apresentadas na Tabela 2.

**Tabela 2. Estatísticas descritivas do desempenho da primeira amostra nos testes DFH, Raven-Escala Colorida e Escala Verbal do WISC-III (QIV).**

Testes	7 – 8 (n=145)		9 –10 (n=137)		11 (n=53)	
	Média	DP	Média	DP	Média	DP
DFH	28,28	5,38	32,98	6,26	34,70	7,45
Raven	23,00	5,58	27,84	4,90	30,48	4,12
QIV	106,00	12,80	110,36	14,50	114,50	9,23

Considerando-se a idade (um indicativo de validade desenvolvimental) verifica-se que a correlação entre esta e o DFH foi de 0,425;  $p < 0,001$ . A idade, portanto, influencia significativamente os escores obtidos no teste. A variância dos resultados no DFH explicada pela idade correspondeu a 18% ( $0,425^2$ ). Quando se considera o teste Raven – Escala Colorida, a correlação entre este e a idade foi de 0,556;  $p < 0,002$  (explicando 31% de variância). Quando se considera o escore total bruto da Escala Verbal do WISC-III, a correlação com idade foi de 0,760 (explicando 58% de variância). O critério de validade desenvolvimental foi, portanto, menos intenso no DFH do que nos testes Raven e Escala Verbal do WISC-III, pelo menos para a primeira amostra.

Para todas as faixas etárias, as correlações de Pearson foram significativas entre o DFH e o teste Raven-Escala Colorida ( $r$  variando entre 0,177 e 0,299). O mesmo padrão se encontra entre o DFH e o QI Verbal do WISC-III ( $r$  variando entre 0,232 e 0,281), com exceção do último grupo etário (11 anos) cujo resultado não foi significativo ( $0,232$ ;  $p = 0,105$ ). Quando se considera a amostra total, a correlação parcial (controlando-se a idade) aponta que o índice de associação do DFH é de 0,204 e de 0,299 com o teste Raven e o QI Verbal respectivamente. No entanto, a correlação entre o teste Raven e o QI Verbal é superior (0,401).

Aplicando-se a fórmula de Steiger (1980) a fim de verificar se as correlações entre o Raven-Escala Colorida e o QI Verbal (WISC-III) são, de fato, superiores às correlações entre o DFH e o Raven e entre o DFH e o QI Verbal, obtêm-se na amostra total, diferenças estatisticamente significativas ( $t = 3,2284$ ;  $gl = 326$ ;  $p = 0,0014$ ). Isto quer dizer que os testes tradicionais de avaliação intelectual como o Raven e a Escala Verbal do WISC-III compartilham mais intensamente componentes cognitivos do que entre o DFH e esses mesmos instrumentos.

A fim de obter informações sobre o grau de associação de cada subteste da escala verbal do WISC-III com o DFH, realizaram-se correlações parciais (controlando-se a idade) com a amostra total. Os resultados encontrados apontam para a existência de correlações baixas, porém significativas, entre o DFH e todos os seis subtestes da Escala Verbal do WISC III, não havendo nenhum subteste com maior associação ao DFH.

### **Segunda amostra: Validade convergente e discriminante**

As estatísticas descritivas (média e desvio padrão) dos testes DFH, R-2 e Bender aplicados na segunda amostra ( $n = 293$ ) são apresentadas na Tabela 3.

**Tabela 3. Estatísticas descritivas da segunda amostra nos testes DFH, R-2 e Bender**

Idades	DFH/Goodenough		DFH/Harris		Goodenough/Harris	
	Média	DP	Média	DP	Média	DP
Testes	5 – 7 (n=98)		8 – 9 (n=115)		10 – 11 (n=80)	
DFH	21,20	6,86	30,46	7,42	33,34	6,37
R-2	12,12	4,89	18,67	4,93	21,02	4,81
Bender	11,60	4,86	5,64	3,69	4,73	3,26

A segunda amostra, mais diversificada em idade e em procedência escolar, mostra forte relação entre a idade e o DFH (0,624;  $p < 0,001$ ). A variância dos resultados no DFH explicada pela idade correspondeu a 39% (0,624<sup>2</sup>). Quando se considerou o Bender, a correlação com a idade foi de -0,635;  $p < 0,001$  (explicando 40% de variância). Quando se considerou o R-2, a correlação com a idade foi de 0,559 (explicação de 31% de variância). O critério de validade desenvolvimental (efeito da idade) foi alto no DFH e no Bender e levemente menor no R-2.

A Tabela 4 mostra, de forma geral, moderadas a baixas correlações entre os testes DFH, R-2 e Bender em todos os grupos etários ( $r$  mínimo de -0,316 e máximo de -0,508). Quando considerada a amostra total, controlando-se a idade, o padrão de correlações fica inalterado, isto é as correlações mantêm-se no nível moderado. Os índices negativos no par DFH e Bender devem-se ao sistema de pontuação do teste Bender (quanto maior é a pontuação da criança menor é a sua capacidade psicomotora). Considerando-se esses índices de correlação, o DFH parece solicitar tanto componentes cognitivos (R-2) quanto componentes psicomotores (Bender).

**Tabela 4. Correlação entre DFH, R-2 e Bender**

Idades	n	DFH x R-2	DFH x Bender	R-2 x Bender
5 – 7	98	0,342*	- 0,508*	- 0,319*
8 – 9	115	0,379*	- 0,391*	- 0,332*
10 – 11	11	0,326*	- 0,425*	- 0,316*
Total	293	0,330*	- 0,396*	- 0,287*

Total = controlando-se a idade

\*n.sig<0,005

Contudo, uma observação mais atenta da Tabela 4 permite vislumbrar três fenômenos:

1. A associação entre o DFH e Bender é maior no grupo de menor idade (5 – 7 anos) do que nos grupos etários maiores.
2. A associação entre o R-2 e Bender mantêm-se inalterada em todos os grupos etários.
3. As associações entre DFH e Bender parecem ser maiores do que a associação entre DFH e R-2, principalmente na faixa etária de 5-7 anos.

Dito de outra forma, o DFH parece solicitar mais os aspectos psicomotores do que os cognitivos nos anos pré-escolares. Essa tendência não se observa no R-2. Uma forma de averiguar a pertinência das observações (1) e (2) é testar a significância estatística das diferenças entre coeficientes de correlações dos diversos grupos etários da mesma amostra aplicando-se a fórmula de Cohen (1988). Na Tabela 5 se observam os valores Z obtidos.

**Tabela 5. Valores  $Z_{obs}$  das diferenças entre as correlações DFH e Bender**

Grupos etários	DFH x Bender	R-2 x Bender
	$Z_{obs}$	$Z_{obs}$
1 x 2	- 0,688	- 0,079
2 x 3	0,284	- 0,039
1 x 3	-0,711	- 0,115

*Nota:* Grupos etários: 1 = 5-7 anos; 2 = 8-9 anos; 3 = 10-11 anos

A Tabela 5 mostra não haver diferenças estatisticamente significativas relacionadas à idade entre os coeficientes de correlação do par DFH/Bender. Portanto, não se confirma o primeiro pressuposto de haver maior associação entre DFH e Bender nas faixas etárias menores do que nos grupos de maior idade. Tampouco há diferenças entre as correlações do R-2 e Bender para os diversos grupos etários.

Para averiguar a pertinência do pressuposto (3), utilizou-se a estatística de Steiger (1980). Especificamente, a fórmula foi aplicada ao trio de correlações [DFH e Bender; DFH e R-2; Bender e R-2] dos grupos etários 1 (5-7 anos) e 3 (10-11 anos). Os resultados mostram não haver diferenças significativas entre os coeficientes de correlação dos três testes em nenhum dos dois grupos etários (Tabela 6).

**Tabela 6. Diferenças entre os três coeficientes de correlação  
 (DFH x Bender / DFH x R-2 / Bender x R-2)**

Idades	t	gl	p
5 - 7	1,5819	5	0,117
10 -11	0,816	77	0,417

## DISCUSSÃO

Quatro pontos de interesse podem ser inferidos a partir dos resultados alcançados. O primeiro deles diz respeito à alta associação entre os sistemas de pontuação (sistemas Wechsler, Goodenough e Harris). No presente estudo, encontrou-se na amostra total, coeficientes entre 0,72 e 0,79. Tais resultados são similares aos obtidos por Alves (1981), que comparou o sistema Goodenough e o sistema Harris encontrando coeficientes entre 0,70 e 0,92. Abou-Jamra e Castilho (1987) informam uma correlação de 0,81 entre esses mesmos sistemas. Em estudo mais recente, Rosa (2008) também



encontra correlações significativas e elevadas ( $r$  variando de 0,80 a 0,90) entre os sistemas de correção Goodenough-Harris e Koppitz, para uma amostra de crianças paulistas de 5 a 11 anos de idade. Portanto, os resultados de validade obtidos no presente estudo não podem ser creditados ao tipo de pontuação utilizada. Pode-se afirmar que é válido empregar um ou outro sistema de análise de desenhos, desde que sua estrutura interna (análise de itens) seja adequada. No caso do DFH-sistema Wechsler, um estudo apontou que, embora haja um alto grau de consistência interna (0,87), o instrumento apresenta problemas de calibração de alguns itens que o compõem (Flores-Mendoza, Abad e Lelé, 2005).

O segundo ponto refere-se à variância explicada pela idade (critério de validade desenvolvimental). Tanto na primeira, quanto na segunda amostra, o DFH mostrou diferenciação de acordo com a idade. Portanto, cumpriu o requisito de validade de desenvolvimento (Anastasi e Urbina, 2000), isto é, instrumentos que visem medir construtos como inteligência devem mostrar variabilidade em função da idade cronológica no período infantil. O fato de ter ocorrido maior proporção de variância explicada pela idade na segunda amostra (39%) do que na primeira (18%), deve-se à amplitude da faixa etária compreendida pelas amostras. Enquanto na primeira, a faixa etária compreendia desde os 7 até os 11 anos; a segunda amostra compreendia idades dos 5 aos 11 anos de idade.

O terceiro ponto refere-se à associação de instrumentos cognitivos (validade convergente) e psicomotores (validade discriminante) com o DFH. O Teste das Matrizes Progressivas de Raven requer raciocínio associado ao fator  $g$  (Jensen, 1998) e componentes viso-espaciais (Colom, Escorial e Rebollo, 2004), razão pela qual é considerada uma medida de inteligência fluida, portanto, menos influenciável pela cultura. A escala verbal do WISC-III é tida como medida de inteligência cristalizada (Jensen, 1998), definida como a capacidade de reproduzir e operar o conhecimento adquirido no ambiente. No estudo de adaptação do WISC-III ao contexto brasileiro (Wechsler, 2002), o índice de associação com o Raven foi de 0,65. No presente estudo, a associação foi de 0,40 (controlando-se a influência da idade). No que se refere ao Teste R-2, este foi construído com base nos princípios do teste Raven e avalia a capacidade edutiva ou educação de relações de material não verbal (Rosa e Alves, 2000). Seu coeficiente de associação com o teste Raven, segundo o manual, foi de 0,60 (sem controle da influência da idade). Esses três instrumentos cognitivos, portanto, associam-se de forma bastante similar. No entanto, a correlação entre estes testes com o DFH diminui drasticamente. Na primeira amostra, obteve-se coeficientes de associação baixos, embora significativos, de 0,204 com o teste Raven e de 0,299 com a Escala Verbal do WISC-III. Na segunda amostra, a correlação com o R-2 se manteve também baixa (0,330), porém significativa. Rosa (2008) encontrou coeficientes bastante semelhantes entre o DFH (sistemas de correção Goodenough-Harris e Koppitz) e o R-2. Esses resultados são próximos, também, ao que consta no manual do DFH (Wechsler, 2003), segundo o qual a associação com o TNVRI (Teste Não Verbal de Raciocínio Infantil) foi de 0,21. Deve-se registrar que Abou-Jamra e Castilho (1987) também informaram um coeficiente de 0,22 entre o sistema Harris e o Teste Raven. Considerando-se tais resultados pode-se inferir que o DFH constitui um instrumento de triagem cognitiva, uma medida intermediária da inteligência (em inglês uma medida “*proxy*”). Isto

---

é, assim como o alcance educacional, a hierarquia ocupacional e o nível sócio-econômico, todas elas relacionadas positivamente ao nível intelectual, e, portanto, consideradas como medidas “proxies” da inteligência, também o DFH é uma medida “proxy” da inteligência. Por exemplo, em estudo recente com crianças mineiras Colom e Flores-Mendoza (2007), encontraram uma associação entre nível socioeconômico e o Raven - Escala Colorida de 0,20 e uma associação entre o nível educacional dos pais e o Raven - Escala-Colorida de 0,19.

O quarto ponto de discussão refere-se à associação encontrada entre o DFH e o Bender. Surpreendentemente, encontraram-se índices levemente superiores aos obtidos entre os testes cognitivos, especialmente nas primeiras faixas etárias, o que provocou a indagação, se haveria maior demanda no DFH de componentes psicomotores do que cognitivos. Essa suspeita estava reforçada pelas informações do próprio manual do DFH em que são mostrados índices de associação de 0,62 com um teste psicomotor e de 0,21 com um teste cognitivo. Contudo, no presente estudo, a estatística de Cohen (1988), mostrou não haver diferenças entre essas demandas nas três faixas etárias. É possível que o tamanho pequeno das amostras em cada grupo etário não tenha permitido detectar as diferenças, razão pela qual se advoga por estudos mais amplos.

Em outra direção, a estatística de Steiger (1980) tampouco mostrou haver diferenças entre os índices de validade convergente (DFH/R-2) e os de validade divergente (DFH/Bender e R-2/Bender). Tais resultados sugerem a ocorrência de uma sobreposição dos componentes psicomotores e cognitivos nos testes DFH, Bender e R-2 e que não desaparece até os 11 anos de idade. Tal conclusão está amparada também em outros estudos. Abou-Jamra e Castilho (1987) informam que Ansbacher na década de 50 encontrou uma correlação de 0,34 entre o DFH e o teste de traçado do Teste de Habilidade Mecânica de McQuarrie aplicados a 100 crianças de 10 anos de idade. No contexto nacional, Bandeira (1992) identificou uma associação de -0,335 entre o Teste Raven e o Bender. Sisto, Noronha e Santos (2006) encontraram uma associação significativa (-0,552) entre o novo sistema brasileiro de pontuação gradual do Bender e o Raven, embora não esteja claro, se em tal estudo ocorreu o controle da idade. Através destes estudos fica evidente que a associação entre os testes cognitivos e o teste psicomotor é característica do desenvolvimento infantil.

Em geral, pode-se afirmar que o DFH é um instrumento útil para detectar de forma preliminar as diferenças individuais no desenvolvimento cognitivo de um determinado grupo (vide resultados de validade desenvolvimental). Contudo, seu uso para efeito de classificações e diagnósticos individuais deve ser usado com cautela e junto com outros testes complementares, haja vista sua modesta associação com testes tradicionais de inteligência.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Abou-Jamra, C.C. & Castilho, M.T.U.R. (1987). Testes de Colúmbia, Raven-Escala Especial, INV e Goodenough. In: M. Ancona-Lopez (Org.). *Avaliação da Inteligência*. (pp. 1-64; vol. II). São Paulo: EPU.
- Alves, I.C.B. (1979). *O Teste Goodenough-Harris em uma população pré-escolar paulistana*. Dissertação de Mestrado. Instituto de Psicologia da Universidade de São Paulo.
- Alves, I.C.B. (1981). O Teste Goodenough-Harris em pré-escolares paulistanos. *Boletim de Psicologia*, XXXIII (80), 40-52.
- Anastasi, A. & Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: Artes Médicas Sul.
- Angelini, A.L., Alves, I.C.B., Custódio, E.M., Duarte, W.F. & Duarte, J.L.M. (1999). *Manual Matrizes Progressivas Coloridas de Raven: Escala Especial*. São Paulo: Centro Editor de Testes e Pesquisas em Psicologia.
- Bandeira, D. (1992). *A contribuição dos testes DFH, Bender e Raven na predição do rendimento escolar na primeira série*. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul.
- Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences*. (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Colom, R., Escorial, S. & Rebollo, I. (2004). Sex differences on the Progressive Matrices are influenced by sex differences on spatial ability. *Personality and Individual Differences*, 37 (6), 1289-1293.
- Colom, R. & Flores-Mendoza, C.E. (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence*, 35 (3), 243-251.
- Flores-Mendoza, C.E. Abad, F.J. & Lelé, J.A. (2005). Análise de itens do Desenho da Figura Humana: Aplicação de TRI. *Psicologia: Teoria & Pesquisa*, 21 (2), 243-254.
- Jensen, A. (1998). *The g factor*. London: Praeger.
- 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.
- Koppitz, E. (1989). *O Teste Gestáltico Bender para Crianças*. Porto Alegre: Artes Médicas.
- Rosa, H.R. (2008). Validade do Desenho da Figura Humana na avaliação de Goodenough-Harris e nos Indicadores Maturacionais de Koppitz em crianças da Cidade de São Paulo. *Boletim de Psicologia*, LVIII, (128), 1-14.
- Rosa, H.R. & Alves I.C.B. (2000). *R-2: Teste Não Verbal de Inteligência para Crianças. Manual*. São Paulo: Editora Vetor.
- Sisto, F.F.; Noronha, A.P.P. & Santos, A.A.A. (2006). *Teste Gestáltico Visomotor de Bender. Sistema de Pontuação Gradual (B-SPG)*. São Paulo: Editora Vetor.
-

- Steiger, J. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87 (2), 245-2251.
- Wechsler, D. (2002). *WISC-III: Escala de Inteligência para Crianças: Manual*. Adaptação para o contexto brasileiro de Vera Lucia M. Figueiredo. São Paulo: Casa do Psicólogo.
- Wechsler, S. & Schelini, P. W. (2006). Bateria de Habilidades Cognitivas Woodcock-Johnson III: Validade de construto. *Psicologia: Teoria & Pesquisa*, 22 (3), 287-296.
- Wechsler, S. M. (2003). *DFH III: O Desenho da Figura Humana: Avaliação do desenvolvimento cognitivo de crianças brasileiras*. Terceira edição. Campinas: LAMP/PUC.

*Recebido em 17/04/09*

*Revisto em 01/09/09*

*Aceito em 05/09/09*