

# COMPARAÇÃO DE MODELOS DA TEORIA DE RESPOSTA AO ITEM (TRI) NA VALIDAÇÃO DE UMA PROVA DE DEPENDÊNCIA-INDEPENDÊNCIA DE CAMPO

*Maria João Vieira – Ray Human Capital  
Rui Bártolo Ribeiro – Instituto Superior de Psicologia Aplicada  
Leandro Almeida – Universidade do Minho  
Ricardo Primi – Universidade São Francisco*

---

## RESUMO

O presente estudo analisa itens de uma prova que avalia o construto dependência-independência de campo – Testes “Padrão” – à luz dos modelos de um, dois e três parâmetros da Teoria de Resposta ao Item (TRI). Utilizou-se uma amostra de 1.918 candidatos ao Curso de Pilotagem Aeronáutica da Academia da Força Aérea Portuguesa. Efetuou-se uma análise fatorial por informação completa com o intuito de verificar as pré-condições da TRI, a qual indicou a existência de dois fatores. Posteriormente, analisou-se cada um dos fatores à luz dos modelos unidimensionais da TRI, revelando-se que a divisão dos dois fatores está relacionada com a dificuldade dos itens, e que os itens mais fáceis são aqueles nos quais a figura-alvo não se encontra em nenhuma das figuras padrão. Concluiu-se que o Subteste 2 se revelou o mais indicado para avaliar o construto, e sugere-se que sejam acrescentados itens fáceis, para que o teste seja mais adequado a sujeitos com baixos níveis de aptidão.

*Palavras-chave:* Teoria da Resposta ao Item (TRI); Dependência-Independência de campo; Validade; Testes.

## COMPARISON OF ITEM RESPONSE THEORY (IRT) MODELS IN THE VALIDATION OF A FIELD DEPENDENCE-INDEPENDENCE TEST

### ABSTRACT

This study analyzes the items of the test Patterns which assess the construct of field dependence-independence - Test. Data were analyzed in the light of one, two and three parameters of Item Response Theory (IRT). We used a sample of 1918 candidates for the instruction at the Portuguese Air Force Pilot Academy. We conducted a full information factor analysis in order to verify the preconditions of IRT that indicated the existence of two factors. Then we examined each factor in light of unidimensional IRT models that revealed that the division of the two factors is related to the item difficulty and that the easiest items are those in which the target figure is not in any of the standard figures. It was concluded that the subtest 2 proved to be the most suitable for assessing the construct, and it is suggested the addition of easier items so that the test is more suitable for individuals with low ability levels.

*Keywords:* Item Response Theory (IRT); Field dependence-independence; Validity; Tests.

---

## INTRODUÇÃO

A Teoria Clássica dos Testes (TCT) apresenta algumas limitações na construção e validação de provas psicológicas. Os parâmetros dos itens que se calculam pela TCT são dependentes da amostra, torna-se dessa forma um item mais fácil ou difícil em função da habilidade dos sujeitos que o respondem. Por outro lado, as habilidades dos sujeitos ficam atreladas ao score bruto num teste específico, perdendo-se a possibilidade de interpretabilidade numa escala geral do construto avaliado. Essa limitação torna-se mais evidente quando há diferentes tipos de testes ou então diferentes versões dos instrumentos avaliando o mesmo construto, situação para a qual a TCT não oferece soluções práticas e simples para a

criação de uma escala com métrica equivalente entre as provas. Há, ainda, o problema da estimativa da precisão que resulta em um valor global para o instrumento sem considerar a influência que a adequação da dificuldade do teste às habilidades das pessoas tem na precisão. Todos esses aspectos, não são devidamente ponderados na TCT, quer seja no nível da pesquisa, quer seja no da prática (Fan, 1998; Hambleton & van der Linden, 1982; Hambleton, Swaminathan, & Rogers, 1991; Pasquali, 2007).

Em alternativa, recorre-se à Teoria da Resposta ao Item (TRI) pelas possibilidades de se ultrapassarem algumas das dificuldades metodológicas presentes na TCT (Wiberg, 2004). A TRI é baseada em modelos matemáticos e métodos estatísticos (Reise, Ainsworth & Haviland, 2005)

nos quais o cálculo da estimativa do traço latente, leva em conta não só das respostas dadas pelos indivíduos, mas também as propriedades dos itens que serviram essa avaliação (Embretson & Reise, 2000). Por exemplo, no quadro de uma avaliação adaptativa, os itens a administrar podem ser escolhidos em função do objetivo de maior ou menor seletividade pretendida com a avaliação e, também, conforme o nível de traço que os indivíduos possuem. Os modelos mais utilizados em testes de inteligência cuja pontuação nos itens é feita de forma dicotômica (acerto ou não acerto) são os Modelos Logísticos de um (ML1), dois (ML2) e três parâmetros (ML3).

Esse modelo de análise tem como alicerces a independência local e a unidimensionalidade dos itens, fortemente associados entre si (Embretson & Reise, 2000; Hambleton, Swaminathan & Rogers, 1991, Pasquali & Primi, 2003). O primeiro postulado requer que, mantidas constantes as aptidões que afetam o teste – à exceção da aptidão que está a ser mensurada ( $\theta$ ), as respostas dadas por um mesmo sujeito a dois itens são independentes (Pasquali & Primi, 2003). Porém, para uma correta aplicação da TRI a um qualquer instrumento de medida, deve também assumir-se a unidimensionalidade do mesmo, ou seja, a existência de uma dimensão principal dominante (Wiberg, 2004). Dessa forma, o postulado da independência local implica o postulado da unidimensionalidade, uma vez que se assume que o  $\theta$  do sujeito e os parâmetros dos itens são os únicos elementos que governam a resposta dada (Pasquali & Primi, 2003).

Hambleton, Swaminathan e Rogers (1991) apontam algumas vantagens da TRI em face da TCT. O cálculo do nível de aptidão do sujeito é independente da amostra de itens, ou, por outras palavras, o seu nível de aptidão é estimado na base dos itens que realiza, cujas dificuldades são controladas de tal forma que a aptidão não irá depender do subconjunto de itens aplicados. Por outro lado, o cálculo dos parâmetros dos itens é independente da amostra de sujeitos, o que não ocorria na TCT, onde os parâmetros dos itens estão dependentes do nível de aptidão dos sujeitos considerados. Finalmente, como corolário dos dois pontos anteriores, na TRI pode efetuar-se o emparelhamento dos itens com a aptidão do sujeito. Na TCT, o mesmo teste era usado para sujeitos com elevada ou baixa aptidão, verificando-se que, se o teste fosse fácil avaliava bem os sujeitos com baixa aptidão, ao passo que se fosse difícil avaliava bem

os sujeitos com aptidões elevadas. A TRI permite, assim, que se avaliem os sujeitos com maior precisão através de itens com um grau de dificuldade adequado ao nível de aptidão do sujeito, nomeadamente recorrendo a testes adaptativos computadorizados.

Neste artigo pretendemos ilustrar as potencialidades de análise da TRI aplicadas aos itens de uma prova psicológica avaliando o estilo cognitivo da dependência-independência de campo (Teste Padrões). Mais concretamente, pretendeu-se efetuar uma análise dos itens empregando dois modelos da TRI que constituem a prova “Padrões”, com o intuito de conhecer o seu grau de dificuldade, poder de discriminação e a probabilidade de acerto ao acaso. Com essa informação, verificou-se se o teste se mostrava adequado na avaliação do construto “dependência-independência de campo”, e foram sugeridas ações de melhoria da prova.

## MÉTODO

### *Amostra*

A amostra deste estudo é constituída por 1.918 candidatos ao Curso de Pilotagem Aeronáutica da Academia da Força Aérea Portuguesa que prestaram provas entre 2004 e 2007. A média de idades dos candidatos é de 18,7 anos, tendo o participante mais novo 17 e o mais velho 23 anos. Os participantes têm, no mínimo, o 12º ano de escolaridade completo, e 87,4% da amostra é do sexo masculino. Trata-se de uma amostra por conveniência, que corresponde à totalidade dos candidatos que prestaram provas nos quatro anos consecutivos.

### *Instrumento*

A prova “Padrões” é um teste de escolha múltipla informatizado, desenvolvido por Eugene Burke em 1995, que avalia o estilo cognitivo da dependência-independência de campo. Esta descreve duas formas distintas de processar informação (Guisande, Páramo, Tinajero & Almeida, 2007; Miller, 2007): os indivíduos dependentes de campo, tendencialmente, percebem o campo como um todo, têm dificuldade em separar a informação do seu contexto e são mais facilmente influenciados por pistas externas; os indivíduos independentes de campo têm tendência para se focar em aspectos isolados, têm mais facilidade em separar a informação essencial do contexto e são mais facilmente influenciados por pistas internas do que externas (Miyake, Witzki & Emerson, 2001).

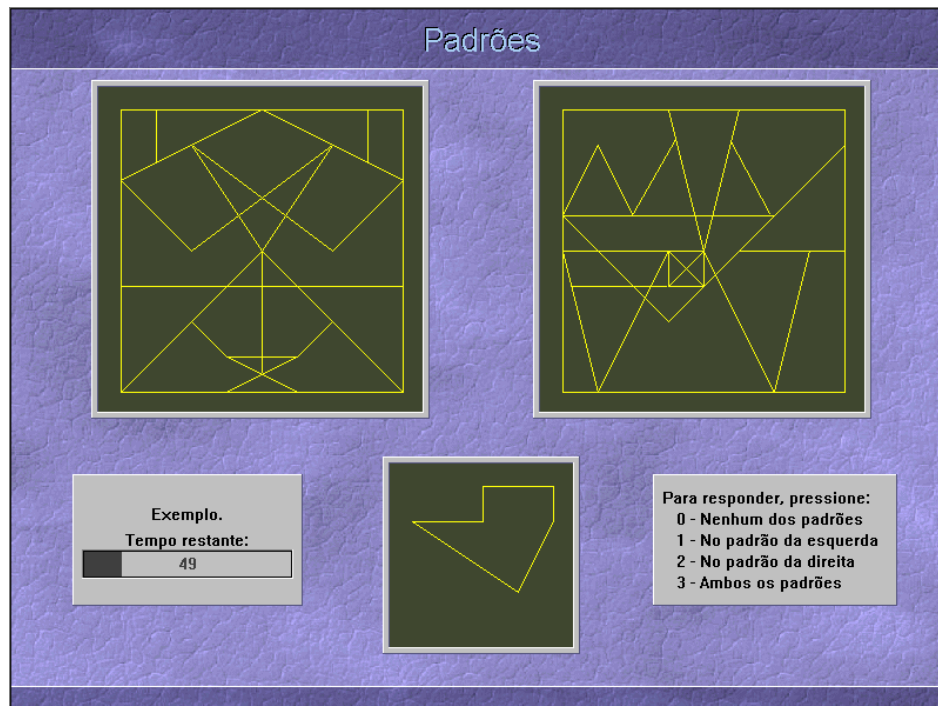


Figura 1. Exemplo de um item do teste “Padrões”

Na realização da prova, o participante tem que identificar se a figura-alvo – estímulo – se encontra em uma, nas duas ou em nenhuma das figuras complexas apresentadas (veja exemplo de um item na Figura 1). A prova é composta por 10 itens, apresentados aos participantes numa sequência preestabelecida, e, para a resolução de cada um desses itens, os candidatos têm 1 minuto. A tarefa de resolução de cada item, tem presente a pressão temporal, porque em janela própria é apresentado, em termos gráficos, o tempo restante em contagem decrescente. Essa barra, nos últimos dez segundos, muda para a cor encarnada. Importante salientar que a pressão do tempo contribui para um aumento da ansiedade do candidato, podendo mesmo constituir-se como um fator não cognitivo limitador do sucesso na tarefa.

Na sua aplicação, solicita-se a cada um dos participantes que identifique uma figura-alvo em dois padrões visualmente desorganizados, podendo estar presente apenas num, nos dois ou em nenhum deles. O candidato deve escolher a sua resposta por meio de um teclado numérico, premindo a tecla zero quando não encontra a figura-alvo em nenhuma das figuras complexas; a tecla um, caso encontre a figura-alvo na figura complexa que se encontra no lado esquerdo; a tecla dois, caso a

figura-alvo esteja apenas na figura complexa do lado direito; e a tecla três, caso se encontre em ambas.

#### *Procedimento*

O teste Padrões fez parte da bateria de provas usada para o ingresso na Academia da Força Aérea Portuguesa, tendo sido aplicada em contexto real de seleção. Em termos de análise estatística dos resultados, os itens foram analisados pelo software TESTFACT (Wilson, Wood & Gibbons, 1991) e BILOG-MG (Zimowski, Muraki, Mislevy & Bock, 1996). Foram efetuadas várias execuções com o objetivo de testar a unidimensionalidade e se proceder à calibração dos itens e dos parâmetros de cada um dos três modelos. Foi também utilizado o Winsteps (Linacre & Wright, 2001), que é um software que permite a realização da análise dos dados à luz do Modelo de Rash ou Modelo Logístico de Um Parâmetro. Importa referir que o Winsteps (Linacre & Wright, 2001) faz a estimação dos parâmetros por máxima verossimilhança, ao passo que o BILOG-MG (Zimowski, Muraki, Mislevy & Bock, 1996) utiliza processos de estimação bayesianos.

## RESULTADOS

Para a verificação do pressuposto da unidimensionalidade dos itens, dada a sua natureza dicotómica, procedeu-se a uma análise prévia dos coeficientes de correlação tetracóricos (Parsons & Hulin, 1982; Primi & Almeida, 1998; 2000; Richaud, 2005) e baseada em informação completa usando a TRI mediante o software TESTFACT (Wilson, Wood & Gibbons, 1991). Essa análise da unidimensionalidade, ao contrário do desejado, permite afirmar que o teste Padrões é composto por dois fatores explicando 18,7% e 5,1% da variância respectivamente. O TESTFACT apresenta a precisão dos fatores usando os erros da estimativa de teta a partir da estimação da distribuição *a posteriori* que resultou nos valores 0,56 e 0,30, respectivamente. O segundo fator, que se vai designar por subtteste 2, é composto pelos itens 1, 2, 3 e 4, ao passo que o primeiro fator (subtteste 1) é composto pelos itens restantes. Importa também referir que os itens 3 e 4 do segundo fator também

possuem carga fatorial superior a 0,20 no primeiro. Salienta-se, ainda, que os itens 3 e 4 são os que apresentam as comunalidades mais baixas, 0,27 e 0,31, respectivamente. Por outro lado, com a atenção focada na unidimensionalidade do teste, é relevante referir que a análise factorial tetracórica, com rotação promax, revela que a correlação entre os dois fatores é de 0,48, o que permite reforçar a ideia de haver uma dimensão hierarquicamente superior a alicerçar cada um dos fatores.

Após a análise fatorial indicando a existência de duas dimensões, realizou-se a estimação dos parâmetros dos itens pelo modelo de Rasch, por meio do Winsteps para cada dimensão em separado. Tal como no estudo realizado por Alonso e colaboradores (2008), optou-se por verificar se os valores dos índices *Infit Mnsq* e *Oufit Mnsq* oscilavam entre os valores considerados aceitáveis por Linacre e Wright (1994). Todos esses índices variam entre 0,88 e 1,07 (cf. Tabela 1), o que nos permite concluir que os itens se ajustam ao modelo (Linacre & Wright, 1994).

Tabela 1 – Índices de ajuste do ML1 ao Subtteste 2

Nº Acertos	N	b	Erro do modelo	Winsteps 3,61								
				Infit		Outfit		PTMEA Corr	Exact Obs%	Match Exp%	Item	
				MNSQ	ZSTD	MNSQ	ZSTD					
1	1.144	1.468	-1,15	0,07	0,92	-2,20	0,88	-2,20	0,53	81,70	78,70	i1
2	782	1.468	0,20	0,06	1,06	2,50	1,07	2,60	0,54	65,70	67,90	i4
3	854	1.468	-0,40	0,06	1,01	0,20	1,00	0,20	0,56	66,20	67,90	i3
4	539	1.468	1,00	0,06	1,01	0,30	0,99	-0,30	0,61	67,70	68,60	i2
Média	829,80	1.468,00	0,00	0,06	1,00	0,20	0,99	0,10		70,40	70,80	
Desvio Padrão	215,70	0,00	0,77	0,00	0,05	1,70	0,07	1,70		6,60	4,60	

De acordo com os dados da Tabela 1, os quatro itens que compõem este subtteste têm graus de dificuldade que oscilam entre -1,15 e 1, o que nos permite afirmar que o espectro de avaliação do Subtteste 2 não é muito amplo, medindo apenas os valores circundantes ao valor médio. Pela análise do mapa item-pessoa, verificou-se que existe um número muito elevado de sujeitos – mais de mil – que tem uma aptidão superior à exigida pelo item mais difícil, pelo que podemos inferir que o teste é

fácil. A informação fornecida pelo ML1, relativa aos itens que compõem este subtteste, levou a crer que estávamos perante um subtteste que, por si só, era insuficiente para estimar a aptidão dos sujeitos.

No Subtteste 1, os valores dos índices de *Infit Mnsq* e *Oufit Mnsq* encontravam-se balizados entre 0,82 e 1,18 (ver Tabela 2), pelo que, uma vez mais, se pode assumir que o modelo se ajustava aos dados, de acordo com Linacre e Wright (1994).

Tabela 2. Índices de ajuste do ML1 ao Subteste 1

Winsteps 3,61													
	Nº Acertos	N	b	Erro do modelo	Infit		Outfit		PTMEA Corr	Exact Obs%	Match Exp%	Item	
					MNSQ	ZSTD	MNSQ	ZSTD					
	2	820	1.638	-0,30	0,06	1,13	5,50	1,18	5,70	0,49	65,90	70,00	i6
	6	494	1.638	0,76	0,06	1,04	1,50	1,08	1,70	0,49	72,20	74,20	i10
	3	498	1.638	0,75	0,06	1,02	0,60	1,05	1,10	0,50	73,40	74,20	i7
	4	990	1.638	-0,83	0,06	1,04	1,60	1,04	1,20	0,55	69,00	70,50	i8
	1	633	1.638	0,29	0,06	0,90	-4,10	0,86	-4,30	0,59	75,10	71,70	i5
	5	939	1;638	-0,67	0,06	0,86	-6,30	0,82	-5,70	0,64	74,70	70,00	i9
Média	729,00	1;638	0,00	0,06	1,00	-0,20	1,01	-0,10			71,70	71,80	
Desvio Padrão	199,30	0,00	0,64	0,00	0,09	3,90	0,12	3,80			3,30	1,80	

Dessa forma, avançou-se para a análise dos parâmetros dos itens. Como se pode verificar na Tabela 2, o item 10 foi o que apresentou maior grau de dificuldade, ao passo que o item 8 se revelou como o mais fácil. Mesmo contendo mais dois itens que o Subteste 2, o 1 apresentou um espectro de dificuldade menor, existindo um par de itens praticamente com o mesmo grau de dificuldade: itens 7 e 10, com dificuldades de 0,75 e 0,76, respectivamente.

A análise do mapa item-pessoa permitiu verificar a existência de cerca de 600 sujeitos com aptidões mais baixas do que a exigida pelo item mais fácil e cerca de duas centenas de candidatos com aptidões mais elevadas do que a exigida pelo item mais difícil (item 10). Estamos, pois, diante de um subteste difícil.

No que diz respeito ao Subteste 2, pelo uso do software BILOG-MG, verificou-se a convergência para o ML2 (maior mudança=0,007<0,01=critério de convergência). No Subteste 1, verificou-se também a convergência para o ML2 (maior mudança=0,005<0,01=critério de convergência). Para verificar a adequação do modelo aos itens recorreremos a uma análise gráfica, mediante a comparação das CCIs empíricas e teóricas. Embora haja algumas discrepâncias visíveis em alguns itens (6 e 8), em geral observou-se uma adequação aceitável.

Com a informação que o ML2 nos fornece, constatou-se que o item que se apresentou com o parâmetro de dificuldade mais elevado, como se pode visualizar na Tabela 3, é o item 10, seguido do item 7; em contraponto, o item mais fácil é o 8, seguido do 9.

Tabela 3. Parâmetros do Subteste 1 (ML2)

Item	Parâmetro a	Parâmetro b
5	0,93	0,55
6	0,35	0,33
7	0,49	1,25
8	0,5	-0,24
9	1,17	-0,04
10	0,43	1,41

Relativamente à dificuldade do subteste, é importante mencionar que, em razão do método de estimação do BILOG-MG, que centra a escala nas habilidades dos sujeitos, é possível afirmar que itens com  $b < 0$  são considerados mais fáceis para a amostra, pois a média da habilidade é igual a zero

( $\theta = 0$ ). Tendo em conta que, dos seis itens que o compõem, apenas dois são considerados fáceis, pode-se concluir que esse foi um teste difícil para os sujeitos. Aliás, esta informação é claramente visível na Função de Informação do Teste, onde o valor máximo de informação é dado em torno do

valor médio de aptidão em que o teste é mais preciso e, sendo a curva ligeiramente assimétrica à direita, pode afirmar-se que nos dá mais informação para valores mais elevados da aptidão.

Já no que diz respeito ao parâmetro  $a$  (Tabela 3), os itens 5 e 9 são os que detêm uma maior discriminação, ainda que esta seja considerada moderada, apresentando os restantes itens um baixo poder discriminativo (Baker, 2001). Esse subteste, por si só, não detém um espectro de medição da aptidão largo, uma vez que está balizado por  $-0,24$  ( $b_8$ ) e  $1,41$  ( $b_{10}$ ); nem tão-pouco se pode afirmar que é um teste equilibrado no que diz respeito à dificuldade dos itens que o compõem, pois como já se verificou, a maioria dos itens exige uma aptidão superior ao seu valor médio para que se possa acertar na resposta.

Para o ML3, e recorrendo uma vez mais ao BILOG-MG (Zimowski, Muraki, Mislevy & Bock, 1996), os dados do Subteste 2 atingiram a convergência para um critério de 0,01, tendo a maior mudança um valor de 0,009 – menor que o critério de convergência. O Subteste 1 também atingiu a convergência para o ML3, tendo sido a

maior mudança verificada de 0,009, quando o critério de convergência assumiu o valor 0,01. Analisando as CCIs empírica e teórica, novamente os itens 6 e 8 apresentam discrepâncias visíveis, embora em geral haja uma correspondência adequada entre as curvas teóricas e empíricas.

Como se pode verificar na Tabela 4, o item que apresentou maior grau de dificuldade foi o 10, seguido pelo item 7; em contraponto, o item 8 revelou-se o mais fácil. Salienta-se que, neste subteste, o item mais fácil exige, para uma correta resposta, um valor de aptidão superior ao valor médio, quer isto dizer que o subteste em análise pode considerar-se difícil. O item 6 apresentou um poder de discriminação baixo, ao passo que os restantes têm o poder de discriminar moderadamente os sujeitos respondentes e o item 9 um poder discriminativo alto. Já no que diz respeito ao parâmetro  $c$ , verificou-se que os sujeitos têm probabilidade de acertar ao acaso entre 5 e 18%, dependendo do item em questão (Tabela 4). De referir que os valores dos três parâmetros do item 9 foram omitidos por não se verificar ajuste do modelo.

Tabela 4. Parâmetros do Subteste 1 (ML3)

Item	Parâmetro a	Parâmetro b	Parâmetro c
5	1,05	0,67	0,1
6	0,48	0,98	0,2
7	0,82	1,41	0,1
8	0,62	0,23	0,2
9	1,39	0,1	0,1
10	0,68	1,6	0,1

Sabemos que quanto maior discriminação um item apresentar, maior quantidade de informação sobre a aptidão estimada dos sujeitos nos fornece, existindo assim uma relação direta entre a discriminação e quantidade de informação. Ora, se um item é muito discriminativo, faz todo o sentido esperar que não tenha uma alta probabilidade de acerto ao acaso por parte dos sujeitos com pouca aptidão. Aliás, na análise deste subteste foi isso mesmo que se verificou: os itens com maior discriminação (item 5 e 9) são os que apresentam menor parâmetro  $c$  (ver Tabela 4). Se existisse algum item com um parâmetro  $c$  de valor muito elevado, seria pertinente retirá-lo, pois apresentaria consequências tão nefastas como: a redução do poder de discriminação do item, e consequentemente a quantidade de informação

fornecida; e o decréscimo efetivo do parâmetro  $b$  (Harvey & Hammer, 1999), uma vez que reduz a informação do item para a estimativa da habilidade.

O teste apresentou um bom nível de informação para uma aptidão maior do que 0, revelando o ponto máximo de informação para um valor de aptidão de 0,4; uma vez que a curva de informação do teste está desviada para a direita, podemos afirmar que está mais bem caracterizada para valores mais elevados de  $\theta$ .

## DISCUSSÃO E CONCLUSÕES

Pretendemos, com este artigo, ilustrar as vantagens de utilização da TRI em relação à TCT em algumas situações concretas da avaliação psicológica, em particular na validade de construto

dos instrumentos usados. Se na TCT se assumem os escores brutos numa prova como medida direta do traço possuído, a TRI demonstra que essa nota bruta calculada pela soma das respostas ao conjunto de itens é, muitas vezes, indevida, pois não estamos a atender à informação específica de cada item. Tomando um teste de estilo cognitivo (dependência *versus* independência de campo), aplicou-se uma análise fatorial por informação completa e concluiu-se que, na realidade, o teste Padrões é composto por dois fatores hierarquicamente organizados, pelo que se procedeu à análise dos fatores à luz dos ML1, ML2 e ML3.

No que diz respeito ao Modelo de *Rasch*, houve adequação aos dois subtestes, podendo retirar-se informação relativa à dificuldade de cada um dos itens que os compõem. Os modelos de dois e três parâmetros também se ajustaram aos dados, informando que há uma variação considerável nos índices de discriminação. É interessante notar que isso não foi impedimento para que o modelo de um parâmetro tivesse um ajuste inadequado aos dados, se considerarmos os valores de corte para os índices *Infit* e *Outfit*. Em geral, há demasiado número de itens com baixa discriminação. Além disso, há uma discrepância quanto à dificuldade dos itens e habilidade da amostra indicando um maior conjunto de itens difíceis para o teste. Isso, aliado ao baixo número de itens, resulta em coeficientes de precisão abaixo do que seria esperado, e esse resultado é obtido qualquer que seja o modelo utilizado para ajustar os dados. O fato de se terem encontrado dois fatores, um com os itens iniciais e outro com os itens finais, pode sugerir que talvez haja uma mudança de estratégia de solução do problema à medida que o sujeito ganha mais experiência no teste. A natureza desses fatores precisa ser mais bem explorada em estudos futuros. Por outro lado, e no sentido da rentabilização destas análises para a construção e validação de testes psicológicos, poder-se-ia analisar o tipo de padrão presente nos estímulos-alvo que caracterizaram os itens fáceis e os itens difíceis. Outra área de estudo futuro passa pela gestão do tempo de realização da prova, já que essa é uma variável importante da resposta.

No sentido de melhorar este teste em concreto, sugere-se o acréscimo de itens que permitam alargar mais o espectro de medição do teta, uma vez que há uma concentração muito maior de itens difíceis. Dessa forma, o teste revela-se pouco adequado, uma vez que não faz uma boa discriminação dos sujeitos com valores mais baixos da aptidão que está a ser medida. Por outro lado,

sabemos que esses sujeitos com pouca aptidão poderiam acertar nos itens por mero acaso. O terceiro parâmetro deste subteste revela que esses sujeitos teriam entre 5 a 18% de probabilidade de acertar nos itens, de acordo com os seus parâmetros *c*.

Pelos resultados encontrados e pelas conclusões retiradas neste estudo, podemos afirmar que a Teoria de Resposta ao Item é uma ferramenta que nos dá acesso a informação pertinente para uma eventual reestruturação da prova. Na realidade, quando se verifica o ajuste dos modelos logísticos de um, dois e três parâmetros, a TRI apresenta claras vantagens na sua aplicação, de entre as quais se destacam a possibilidade de conhecer a contribuição de cada um dos itens para o resultado final e a comparação de testes compostos por diferentes itens (Embretson & Reise, 2000).

## REFERÊNCIAS

- Alonso, J., Olivares, J., Ciudad, A., Manresa, J., Casado, A. & Gilaberte, I. (2008). Desarrollo y validación de la versión corta de la Escala de Funcionamiento Social en esquizofrenia para su uso en la práctica clínica. *Actas Espanolas de Psiquiatria*, 36(2), 102-110.
- Baker, F. B. (2001). *The basics of Item Response Theory*. Washington, DC: ERIC.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for psychologists*. Hillsdale, NJ: Lawrence Erlbaum.
- Fan, X. (1998). Item Response Theory and Classical Test Theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58, 357-381.
- Guisande, M., Páramo, M., Tinajero, C. & Almeida, L. (2007). Field dependence-independence (FDI) cognitive style: An analysis of attentional functioning. *Psicothema*, 19(4), 572-577.
- Hambleton, R. K. & van der Linden, W. J. (1982). Advances in Item Response Theory and applications: An introduction. *Applied Psychological Measurement*, 6, 373-378.
- Hambleton, R., Swaminathan, H. & Rogers, J. (1991). *Fundamentals of Item Response Theory*. New York: Sage.
- Harvey, R. & Hammer, A. (1999). Item Response Theory. *The Counseling Psychologist*, 27, 353-383.

- Linacre, J. & Wright, B. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(2), 370.
- Linacre, J. & Wright, B. (2001). *Winsteps* (Versão 3.61.1) [Computer software]. Chicago: Mesa Press.
- Miller, A. (2007). Creativity and cognitive style: The relationship between field-dependence-independence, expected evaluation, and creative performance. *Psychology of Aesthetics, Creativity, and the Arts*, 1(4), 243-246.
- Miyake, A., Witzki, A. & Emerson, M. (2001). Field dependence-independence from a working memory perspective: A dual-task investigation of the Hidden Figures Test. *Memory*, 9(4-6), 445-457.
- Parsons, C. & Hulin, C. (1982, December). An empirical comparison of Item Response Theory and hierarchical factor analysis in applications to the measurement of job satisfaction. *Journal of Applied Psychology*, 67(6), 826-834.
- Pasquali, L. (2007). *Teoria de Resposta ao Item*. Brasília, DF: Laboratório de Pesquisa em Avaliação e Medida.
- Pasquali, L. & Primi, R. (2003). Fundamentos da Teoria de Resposta ao Item – TRI. *Avaliação Psicológica*, 2, 99-110.
- Primi, R. & Almeida, L. (1998). Considerações sobre a análise factorial de itens com resposta dicotômica. *Psicologia: Teoria, Investigação e Prática*, 3(2), 225-234.
- Primi, R. & Almeida, L. (2000). Estudo de validação da Bateria de Provas de Raciocínio (BPR-5). *Psicologia: Teoria, Investigação e Prática*, 16(2), 165-173.
- Reise, S., Ainsworth, A. & Haviland, M. (2005). Item Response Theory. *Current Directions in Psychological Science*, 14(2), 95-101.
- Richaud, M. (2005). Desarrollos del analisis factorial para el estudio de item dicotomicos y ordinales. *Interdisciplinaria Revista de Psicología y Ciencias Afines*, 22(2), 237-251.
- Wiberg, M. (2004). *Classical test theory vs. Item Response Theory*. Retrieved May 10, 2007, from <http://www.umu.se/edmeas//publikationer/pdf/EM%20no%2050.pdf>
- Wilson, D., Wood, R. & Gibbons, R. (1991). *TESTFACT: test scoring, item statistics, and item factor analysis* (Versão 4.0) [Computer software]. Chicago: Scientific Software International (SSI).
- Zimowski, M., Muraki, E., Mislevy, R. & Bock, R. (1996). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items* (Versão 3.0) [Computer software]. Chicago: Scientific Software International (SSI).

Recebido em julho de 2010  
Reformulado em novembro de 2010  
Aceito em maio de 2011