
TEORIA DE RESPOSTA AO ITEM (TRI): CONCEITOS ELEMENTARES DOS MODELOS PARA ITENS DICOTÔMICOS

GLEIBER COUTO

Laboratório de Avaliação, Medidas e Instrumentação em Psicologia – LAMI
Universidade Federal de Goiás Campus de Catalão - GO- BRASIL

RICARDO PRIMI

Universidade São Francisco - SP - BRASIL

RESUMO

Neste artigo são apresentados os conceitos elementares referentes à medida psicológica sob o ponto de vista da Psicometria Moderna, também chamada de Teoria de Resposta ao Item. São abordados os principais modelos de estimação e descritos os parâmetros de discriminação, dificuldade e probabilidade de resposta correta ao acaso, assim como são analisadas várias implicações relativas à utilização dos diversos métodos de estudo. São discutidas também as características dos procedimentos de análise usados para estimação das curvas características dos itens, curvas características dos testes, curvas características do sujeito, função de informação do item e do teste, erro padrão de medida, definindo os parâmetros da medida.

Palavras-chave: *Psicometria, teoria de resposta ao item, testes psicológicos, avaliação psicológica.*

ABSTRACT

ITEM RESPONSE THEORY (ITR): ELEMENTARY CONCEPTS FOR DICOTOMIC ITEMS MODELS

In this article are presented the elementary concepts relative to psychological measure under the point of view of Modern Psychometry, also named Item Response Theory. The main estimative models are presented and also are described the discrimination parameters, difficulty and probability of random correct answer, as well as several relative implications are analyzed to the use of the several study methods. They are also discussed the characteristics of the analysis procedures used for estimate of the characteristic curves of items, characteristic curves of tests, characteristic curves of subject, function of information of item and of test, standard error of measurement, defining the parameters of the measure.

Key words: *Psychometry, item response theory, psychological tests, psychological assessment.*

INTRODUÇÃO

Atualmente o uso de instrumentos psicológicos, em todas as esferas de atuação profissional do psicólogo, vem sendo ampliado e se tornando corrente. Tal ampliação pode ser notada, de modo especial, no processo de avaliação psicológica, no qual, o uso de tais procedimentos, se não é obrigatório, é bastante freqüente. Desta forma, espera-se cada vez mais que os instrumentos apresentem altos padrões de qualidade (Anastasi e Urbina, 2000).

A qualidade dos instrumentos psicológicos é objeto de estudo da psicometria, estando fortemente associada aos testes e escalas psicométricas. Conseqüentemente é sua atribuição uma constante revisão dos procedimentos usados na estimação das propriedades psicométricas dos testes psicológicos em geral. Não obstante, os procedimentos baseados no modelo clássico da psicometria apresentam limitações que se refletem na qualidade dos testes, de maneira que foram propostas soluções para tais fragilidades. O aprimoramento de tais procedimentos culminou no modelo conhecido atualmente como *Teoria de Resposta ao Item* ou simplesmente TRI (Pasquali, 1997).

Muñiz (1990), Embretson e Reise (2000) apresentam a TRI como o enfoque da teoria dos testes psicológicos que tem como proposta resolver problemas apresentados pelo modelo clássico, a saber, (A) a dependência que a medida apresenta em relação ao tipo de teste usado, (B) a amostra da população usada para a estimação dos parâmetros e (C) a consideração do escore total como referência de medida. Apesar disso, a TRI não veio para substituir o modelo clássico, constituindo-se como um coadjuvante como veremos a seguir.

O modelo não é novo, mostrando suas origens no trabalho de Richardson (1936), Lawley (1943), Tucker (1946), Lord (1952,1953), tendo uma rápida expansão nos anos 60 com o trabalho de Rasch e o desenvolvimento dos computadores pessoais. Sua principal contribuição do ponto de vista teórico é a invariância dos parâmetros de medida, além de apresentar inovações técnicas como as funções de informação dos itens e do teste; medidas mais refinadas dos *erros padrão de medida* (EPM), que permitem observar suas variações ao longo da escala; a possibilidade de atribuição de significado psicológico para interpretação de escalas baseadas nas respostas aos itens.

Os principais pressupostos teóricos do modelo podem ser descritos por meio da Curva Característica dos Itens (CCI) (Figura 1). Nos modelos da TRI é assumida uma relação entre o valor do traço latente (fenômeno psicológico) medido por um teste, representado pela letra theta (θ) e a probabilidade de resposta correta de um sujeito nos itens que compõem esse teste. Essa relação é expressa por uma função conhecida como $P_i(\theta)$, que significa a probabilidade de acerto de um item i , dado um valor de (θ). A função $P_i(\theta)$ é representada no eixo das ordenadas e compõe a representação gráfica da Curva Característica dos Itens (CCI), e no eixo das abscissas, é representado o valor de (θ) (Muniz, 1990; Baker, 2001).

A CCI informa as diferentes probabilidades de acerto que diversos sujeitos com valores diferentes de variáveis latentes (θ) apresentam. O valor de (θ) varia de $-\infty$ a $+\infty$ e, por sua vez, o valor de $P(\theta)$ varia de 0 a 1. A CCI pode informar ainda os parâmetros psicométricos dos itens, conforme o modelo utilizado, a saber, a capacidade de discriminação do item, sua dificuldade e a probabilidade de acerto ao acaso (Pasquali, 1996, 2000).

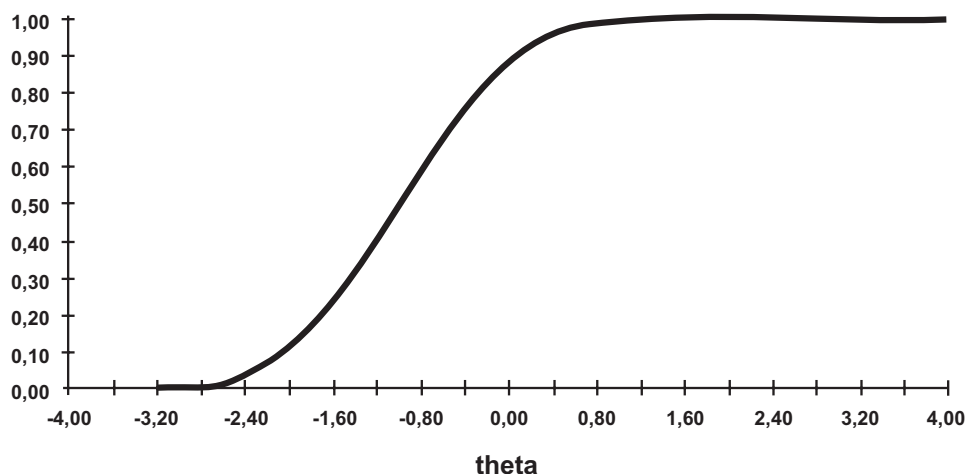


Figura 1. Curva Característica do Item

Os modelos da TRI variam conforme os parâmetros dos itens considerados para avaliação. O primeiro deles é conhecido como a e é denominado índice de discriminação do item, seu valor é dado pela inclinação da CCI em relação ao eixo das abscissas, o valor do ângulo formado por essa inclinação é proporcional ao valor de a que será tanto maior quanto maior for o ângulo. O índice de discriminação na TRI mantém significado correspondente ao da Teoria Clássica, ou seja, a capacidade de discriminar pequenas diferenças no traço latente (Garcia, Maranon, Falcon e Costas, 2001).

O próximo parâmetro é conhecido como b e é denominado índice de dificuldade do item, seu valor é dado pelo ponto, na reta, perpendicular ao eixo das abscissas, e representa o valor de (θ) quando a probabilidade de o sujeito acertar ao item é de 50%, quando $(c = 0)$. A característica mais importante desse parâmetro é que ele se encontra na mesma escala que a variável latente (θ) , isso permite que a dificuldade dos itens possa ser interpretada em termos de variações padronizadas na habilidade dos sujeitos (Garcia et al., 2001).

O parâmetro conhecido como c é denominado probabilidade de acerto ao acaso, seu valor é dado por $P(\theta) \neq 0$ quando $\theta = -\infty$, ou seja, quando a habilidade do sujeito tende à menor representação possível de (θ) e ainda assim a probabilidade de acerto se mantém diferente de zero. Isso significa que o sujeito tem uma habilidade menor que a exigida para a execução adequada do item e mesmo assim sua probabilidade de acerto é diferente de zero.

A curva característica do item (CCI) foi definida conceitualmente como uma propriedade típica da TRI e sua elaboração depende da especificação dos parâmetros dos itens a serem avaliados e a opção por uma função matemática para expressar as curvas dos itens. Em outras palavras, a forma como as curvas se apresentam nos gráficos ilustram os parâmetros dos itens nelas representados. Então, se forem observadas as curvas que se dispõem da esquerda para a direita num continuum, no eixo das abscissas, maior o valor de b conforme se deslocam para a direita; por exemplo, na Figura 2 o item 3 é o mais difícil, enquanto o item 4 é o mais fácil. Quanto mais inclinadas as curvas estiverem

em relação ao eixo das abscissas de modo a formarem um ângulo reto maior o valor de a ; por exemplo, na Figura 2 o item 2 é o mais discriminativo, ao passo que o item 1 é o menos discriminativo. Já o valor de c é representado na origem da curva em relação ao eixo das ordenadas, sua magnitude é equivalente ao valor do deslocamento desse ponto em relação ao valor zero, por exemplo, na Figura 2 o item 4 apresenta uma probabilidade de acerto ao acaso de 0,2 (Hambleton, 1990).

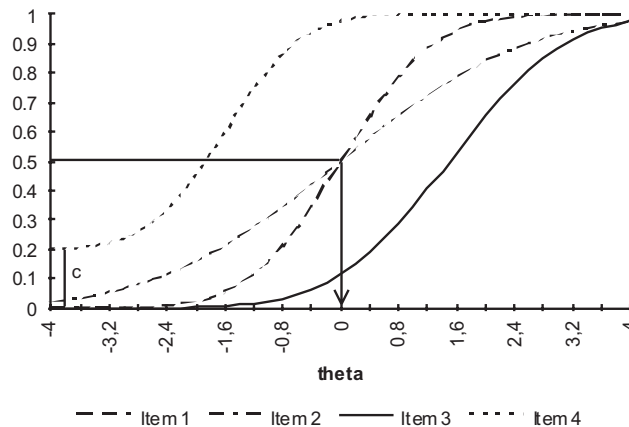


Figura 2. CCI apresentando os parâmetros dos itens

Alguns aspectos precisam ser observados quanto à possibilidade de aplicação dos modelos da TRI aos dados, a saber, a satisfação de dois critérios conhecidos como critério de unidimensionalidade e de independência local. Ao assumir que existe uma relação entre as respostas dos sujeitos e a dimensão latente a ser avaliada (θ), uma exigência para utilização do modelo é que a probabilidade de acerto dependa unicamente, ou pelo menos principalmente, do valor de (θ). Para satisfazer essa condição deve-se verificar se os itens usados para avaliar (θ) medem apenas a dimensão descrita por (θ).

Teoricamente a unidimensionalidade perfeita pode ser encontrada em circunstâncias nas quais a resposta correta aos itens reunidos para avaliar uma determinada característica psicológica é determinada apenas por essa característica. Se a resposta correta aos itens depende ou sofre influência de outras dimensões psicológicas, então não existe unidimensionalidade. Como numa ciência do comportamento essas dimensões são inferidas a partir do comportamento observável do sujeito e é sabido que os comportamentos humanos são multi-causados, o estabelecimento dessa condição estaria de antemão comprometido. Então, para satisfazer o critério da unidimensionalidade a regra adotada em psicometria é a existência de uma dimensão predominante dentre as várias que influenciam o comportamento (Pasquali, 1996, 2000).

Dentre os métodos para se verificar a unidimensionalidade a análise fatorial é o mais difundido e possibilita verificar qual o número mínimo de fatores que pode explicar a quantidade da variância observada. Como não existe unidimensionalidade perfeita, ela é tratada nos modelos da TRI como uma questão de graus, sendo seu índice expresso pelo quociente entre a variância explicada pelo

primeiro fator e a explicada pelo segundo fator, ou seja, quanto maior o valor desse quociente indica o quanto a variância explicada pelo primeiro fator é superior à explicada pelo segundo fator, sugerindo maior unidimensionalidade (Muñiz, 1990; Embretsom e Reise, 2000). Segundo Lord (1980), outro critério prático para se verificar a unidimensionalidade é, na matriz de correlações tetracóricas, se extrair as raízes latentes entre os itens com as comunalidades na linha diagonal. Se a primeira raiz é notadamente superior à segunda, e esta não apresenta diferença significativa entre as outras, então os itens podem ser considerados aproximadamente unidimensionais.

O critério de independência local menciona que a resposta do sujeito a um item não pode ter influência em sua resposta a outros itens, ou seja, existe independência nas respostas entre os itens. Esse critério está diretamente ligado à unidimensionalidade. Se todos os itens medem uma mesma dimensão e a posição do sujeito nessa dimensão não muda (varia) enquanto ele responde ao teste, então as respostas aos itens são estatisticamente independentes, isto é, a probabilidade de acertar um item não depende do acerto aos itens prévios. E, portanto os acertos ou erros dos sujeitos em cada item são independentes entre si, só dependem do theta que é constante durante a aplicação.

OS MODELOS DA TEORIA DE RESPOSTA AO ITEM

Existem vários modelos de TRI dependendo do tipo de função matemática adotada e dos parâmetros dos itens que se queira investigar. Dois tipos de função podem ser encontrados na literatura: as funções logísticas e as de curva normal acumulada. Juntas elas produzem pelo menos seis modelos, sendo possível avaliar até três parâmetros para cada função. Serão abordados neste trabalho apenas os modelos de função logística, pois permitem melhor tratamento matemático e também são mais frequentes na literatura especializada (Muñiz, 1990).

O primeiro modelo é o logístico de um parâmetro que é conhecido como modelo de Rasch, uma referência ao nome de seu idealizador. Trata-se do modelo mais difundido devido a sua parcimônia de medida e da simplicidade de sua lógica. Esse modelo é representado pela função logística de um parâmetro, a qual considera que as respostas de um sujeito a um conjunto de itens dependem apenas de sua habilidade e da dificuldade dos respectivos itens (Baker, 2001; Linacre e Wright, 2002). Sua expressão matemática é:

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}}, \text{ onde,}$$

$P_i(\theta)$, probabilidade de se acertar ao item i dado um determinado valor de (θ) ,

θ , valor do traço latente ou variável que se estiver medindo,

b_i , índice de dificuldade do item i ,

e , base dos logaritmos neperianos que vale (2,72),

D , constante de aproximação aos valores da curva normal acumulada (1,7).

O modelo é bastante simples e conforme a organização de suas variáveis interpreta-se que, conhecendo a dificuldade do item e a habilidade do sujeito, é possível prever qual é a probabilidade desse sujeito acertar o item. O valor de b é dado pelo valor de θ no qual a probabilidade de acertar o item é de 50%.

O modelo logístico de dois parâmetros mantém todas as características do modelo de Rasch. Acrescenta, por sua vez, a estimação do parâmetro discriminação do item. Pode-se assumir que o conceito de discriminação do item é o mesmo usado na Teoria Clássica. Sua expressão matemática é,

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}, \text{ onde,}$$

As variáveis são as mesmas descritas na equação anterior com a diferença que se acrescenta o valor de a que representa o poder discriminativo do item. Uma relação importante entre os parâmetros θ e a é que a discriminação varia em função da habilidade do sujeito. Os itens são mais discriminativos na situação em que a dificuldade dos itens coincide com a habilidade dos sujeitos, nestes casos, o valor de b coincide com o ponto de inflexão da curva, ou seja, o ponto em que a curva muda de direção, tornando-se mais inclinada em relação ao eixo das ordenadas (Muñiz, 1990).

O modelo logístico de três parâmetros mantém todas as características dos modelos de um e dois parâmetros, porém acrescenta a possibilidade de se avaliar a probabilidade de o sujeito acertar o item ao acaso, ou seja, sem que tenha habilidade suficiente. Sua expressão matemática é:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}, \text{ onde,}$$

As variáveis dessa equação são as mesmas da equação anterior com o acréscimo da variável c que representa a probabilidade de acerto ao acaso. O valor de c coincide com o valor de $P_i(\theta)$ para um valor de $\theta = -\infty$. Uma consideração importante, quando se trata dos modelos de um ou dois parâmetros é que para $\theta = b$, $P(\theta) = 0,50$ no modelo de três parâmetros temos para $\theta = b$, $P(\theta) = \frac{1+c}{2}$.

Segundo Muñiz (1990) alguns autores propõem um modelo logístico de quatro parâmetros, que visa controlar circunstâncias aleatórias relacionadas com falhas do construtor no momento da elaboração dos itens, isso faz com que um sujeito com grande competência falhe na resolução do item. Sua expressão matemática é:

$$P_i(\theta) = c_i + (Y_i - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}, \text{ onde,}$$

Todas as variáveis são as mesmas, sendo Y uma tentativa de representar as circunstâncias de falha na construção dos itens e adquire valores pouco menores que um. Até o momento existem poucas pesquisas sobre esse modelo e aparentemente não existe nenhuma vantagem dele em relação ao modelo de três parâmetros.

ESTIMAÇÃO DOS PARÂMETROS

Tendo em vista os principais modelos da TRI, qual critério de escolha o pesquisador deve adotar para analisar os seus dados? Um critério importante a ser considerado refere-se à adequação dos dados ou como é comumente chamado de “ajuste” do modelo aos dados. O ajuste do modelo aos dados escolhido deve ser comprovado pelos parâmetros estimados e comumente segue alguns passos bem definidos (Baker, 2001; Wright e Stone, 2004).

O primeiro passo é comum ao processo de construção de instrumentos de medida em Psicologia de um modo geral. Trata-se da definição rigorosa da dimensão que se pretende avaliar, seguida da elaboração de itens que representem adequadamente essa dimensão. Existem regras a seguir na construção dos itens, porém Muñiz (1990) adverte que o trabalho de construção dos itens é semelhante ao processo de se escrever uma novela, se seguir rigorosamente as regras levasse a boas novelas, então todos seriam excelentes novelistas.

O segundo passo é a aplicação dos itens a uma amostra representativa da população para a qual se constrói o instrumento, com a finalidade de estimar os parâmetros da psicometria clássica e verificar a unidimensionalidade dos itens. Acredita-se que nessa fase já se torne possível vislumbrar qual dos modelos melhor se ajustam aos dados, por exemplo, se a discriminação dos itens não é constante, supõe-se que os dados se ajustam melhor ao modelo de dois parâmetros. Se houver probabilidade de acerto ao acaso, os dados se ajustam melhor ao modelo de três parâmetros. À parte das suposições possíveis nessa fase, os árbitros do processo devem ser os valores dos índices de ajuste que representam a adequação do modelo para analisar os dados. Mas, esse processo não tem regras rígidas, por exemplo, em itens de múltipla escolha geralmente se observam acertos ao acaso, sugerindo-se o uso do modelo de três parâmetros. Entretanto ao se analisar os dados, usando o modelo de um parâmetro, que supõe $c=0$, o resíduo gerado pode ser tão inexpressivo que o uso desse modelo poderá ser apropriado nesse caso.

A partir das respostas dos sujeitos aos itens se inicia a estimação dos parâmetros. Um dos métodos usados é chamado Máxima Verossimilhança no qual os valores são aqueles que maximizam a probabilidade de ocorrência dos dados. Os valores estimados são aqueles que se fazem mais plausíveis para os dados obtidos. O problema da estimação dos parâmetros dos itens na TRI é que, tanto o valor de θ dos sujeitos quanto os valores dos parâmetros dos itens são desconhecidos, conhece-se apenas as respostas dadas aos itens. Como os parâmetros são desconhecidos, é necessário realizar um processo iterativo, como, por exemplo, assumir valores iniciais hipotéticos para os parâmetros dos itens (geralmente derivados de índices da psicometria clássica), estimar as habilidades dos sujeitos, considerar esses novos valores provisórios para re-estimar os parâmetros dos itens de maneira um pouco mais acurada e assim sucessiva e interativamente até que não se consiga melhorar mais os ajustes das curvas teóricas aos dados empíricos.

Os procedimentos de estimação para esse método podem assumir duas formas distintas. A primeira consiste em fixar valores para os parâmetros dos itens e se realizar sucessivas estimações para diferentes valores possíveis de θ , até que o valor que melhor explique a ocorrência do padrão empírico de respostas seja encontrado, são testadas todas as possibilidades. Esta forma é conhecida por estimação condicional como uma referência à condição do conhecimento prévio dos parâmetros

dos itens. A segunda forma consiste em se calcular ao mesmo tempo tanto os parâmetros dos itens quanto o θ de cada sujeito e é conhecida como estimação conjunta. Primeiramente, se estima os valores de θ para cada sujeito, supondo os parâmetros dos itens tal como na estimação condicional, indicando um valor inicial, após os valores de θ conhecidos se retorna calculando os parâmetros dos itens para aqueles valores. A diferença entre os procedimentos é que, no primeiro caso, se estimam as habilidades de todos os sujeitos para depois se corrigir o valor dos parâmetros e no segundo caso, as estimações da habilidade e dos parâmetros dos itens são feitas concomitantemente (Muñiz, 1990).

Para ilustrar os passos usados pelo método de Máxima Verossimilhança, suponhamos que X sujeitos tenham respondido a Y itens que compõem um teste qualquer. São desconhecidos tanto os parâmetros dos itens como as habilidades dos respectivos sujeitos, então o primeiro passo consiste em separar os sujeitos em grupos ao longo de uma escala de habilidade hipotética, cada grupo tem Z sujeitos de habilidades iguais. A probabilidade de os sujeitos de cada grupo responderem adequadamente a um item específico será dada pelo quociente entre o número de sujeitos que realmente acertaram ao item e o número total de sujeitos daquele grupo. Dessa forma as probabilidades de acerto em cada nível de habilidade ao longo da escala podem ser calculadas, isto é, tem-se uma curva empírica para cada item. A partir disso tenta-se manipular os parâmetros do item, produzindo uma curva teórica que mais se aproxime da empírica. O processo de estimação dos parâmetros se encerra quando os valores estimados convergirem, ou seja, quando a partir de n interações não se consegue produzir mais melhorias na reprodução dos dados empíricos por meio das variações nos valores dos parâmetros dos itens (Wright e Stone, 2004; Baker, 2001; Muñiz, 1990).

Após a estimação dos parâmetros do modelo os valores encontrados devem ser confrontados com os resultados empíricos, ou seja, as respostas dos sujeitos, e verificar se existem diferenças estatisticamente significativas. Existem vários procedimentos estatísticos usados para a comprovação do ajuste do modelo aos dados, mas nenhum é totalmente satisfatório, o que acaba por se tornar o ponto fraco da TRI no seu momento atual de desenvolvimento (Muñiz, 1990; Baker, 2001; Wright e Stone, 2004).

A demonstração do ajuste do modelo aos dados pode ser feita por vários caminhos o primeiro a ser abordado é conhecido como Qui-quadrado (χ^2) cuja fórmula para avaliar item por item individualmente é a seguinte:

$$\chi^2 = \sum_{j=1}^k n_j \frac{[P(\theta_j) - P_e(\theta_j)]^2}{[P(\theta_j)][1 - P(\theta_j)]}, \text{ onde}$$

k , número de categorias em que se divide θ

n_j , número de sujeitos dentro de cada categoria.

$P(\theta_j)$, valor da CCI dado pela fórmula do modelo com os parâmetros estimados para a categoria j .

$P_e(\theta_j)$, proporção de sujeitos que empiricamente superam o item para uma categoria determinada j .

χ^2 , se distribui com $k-1$ graus de liberdade.

A fórmula do χ^2 para o cálculo de vários itens é exatamente igual à primeira, bastando executar um somatório de cada resultado para os itens individuais que compõem o teste. Um questionamento freqüente versa sobre a quantidade de categorias que θ pode assumir. Não existe uma resposta definitiva, apenas que o mais comum entre os pesquisadores é se adotar 10 ou 15.

A próxima forma de se demonstrar o ajuste é conhecida como análise dos resíduos. Estes entendidos como a diferença entre o padrão de respostas esperado e o padrão encontrado para um sujeito com determinada habilidade, respondendo a um conjunto de itens com parâmetros já estimados. Assemelha-se muito com o procedimento anterior e é dado pela seguinte fórmula,

$$RE = \frac{P(\theta_j) - P_e(\theta_j)}{\sqrt{P(\theta_j)Q(\theta_j)/n_j}}, \text{ onde,}$$

n_j , número de sujeitos dentro da categoria j .

$P(\theta_j)$, valor da CCI para o nível θ_j .

$P_e(\theta_j)$, proporção empírica de sujeitos dentro de uma categoria dada j que superam o item.

$Q(\theta_j) = 1 - P(\theta_j)$, sujeitos dentro de uma categoria j que erram o item.

À medida que os valores dos resíduos se distanciam de zero, pior será o ajuste do modelo. É comum o estabelecimento de parâmetros arbitrários, por exemplo, aceitar variações entre -2 e 2 como parâmetros aceitáveis de desajuste. O procedimento mais freqüente de análise de ajuste é o χ^2 , que apresenta problemas, quando se trata dos modelos de dois e três parâmetros usado com amostras reduzidas. Esse procedimento deve ser acrescido de outros, por exemplo, análise dos resíduos para estimativas de ajuste complementares (Soares e Pereira, 2002).

A INVARIÂNCIA DOS PARÂMETROS

Uma das vantagens dos modelos da TRI em relação aos modelos da Psicometria Clássica seria a independência dos resultados em relação à amostra que se usou para estimação dos parâmetros e a independência da medida em relação ao instrumento utilizado para procedê-la. Se todos os problemas com a estimação são superados, então o modelo deve ser capaz de alcançar a invariância dos parâmetros.

Quando se utilizam amostras adequadas da população, adequadas no sentido de ser grande o suficiente para representarem as variações possíveis de θ , então se pode dizer que foi encontrada a situação ideal para estimação dos parâmetros. Mas, mesmo quando as amostras são reduzidas e compostas por partes distintas da população para a qual se pretende construir o teste, situação freqüentemente vivenciada pelos construtores de testes, as estimativas demonstram que os parâmetros se mantêm invariantes. Por exemplo, se os parâmetros são estimados com uma amostra da população considerada superior na magnitude do respectivo traço os valores estimados corresponderão à cauda superior da CCI, sendo que o modelo ao estimar os parâmetros deduz o restante inferior da curva para aquela amostra. O mesmo se aplica no caso de a amostra ser constituída pela porção

inferior da população, só que neste caso o modelo deduz a porção superior da curva (Baker, 2001). Evidentemente que, nessas situações, o erro da estimativa será maior, mas em muitos casos, mesmo sendo maior, não chega a ser grande o suficiente ao ponto de invalidar as estimativas.

Dessa forma pode-se dizer que os parâmetros são independentes da amostra utilizada para estimação dos parâmetros. Uma forma de se demonstrar essa invariância é estabelecer uma correlação, por exemplo, a correlação de Pearson entre os valores dos parâmetros obtidos em uma amostra e outra, quanto melhor for essa correlação, ou seja, quanto mais o gráfico de dispersão produzido pelos dados se aproximar de uma reta, mais invariantes são os parâmetros. Ainda, para demonstrar a independência da medida em relação ao instrumento recorre-se também a uma correlação, só que dessa vez entre os valores de θ medidos por dois conjuntos de itens diferentes. Devemos considerar que os valores de θ não serão os mesmos em uma situação e outra se as notas não estiverem equalizadas, pois não existe uma única métrica para θ , mas basta que se estabeleça uma relação linear diretamente proporcional entre as estimações (Baker, 2001; Wright e Stone, 2004). A Figura 3 representa as relações explicadas sobre a independência dos parâmetros com relação à população usada para estimá-los. As curvas normais A e B abaixo do eixo das abscissas representam duas populações com diferentes médias de habilidades que responderam ao item a representado pela curva característica do item. Pode-se notar na faixa sombreada que os indivíduos com habilidade em torno da dificuldade do item representam respectivamente o extrato de maior habilidade da população A e menor habilidade da população B. A dificuldade do item, tal como estimada pela TRI, permanece a mesma, independente da habilidade da amostra utilizada na sua estimação.

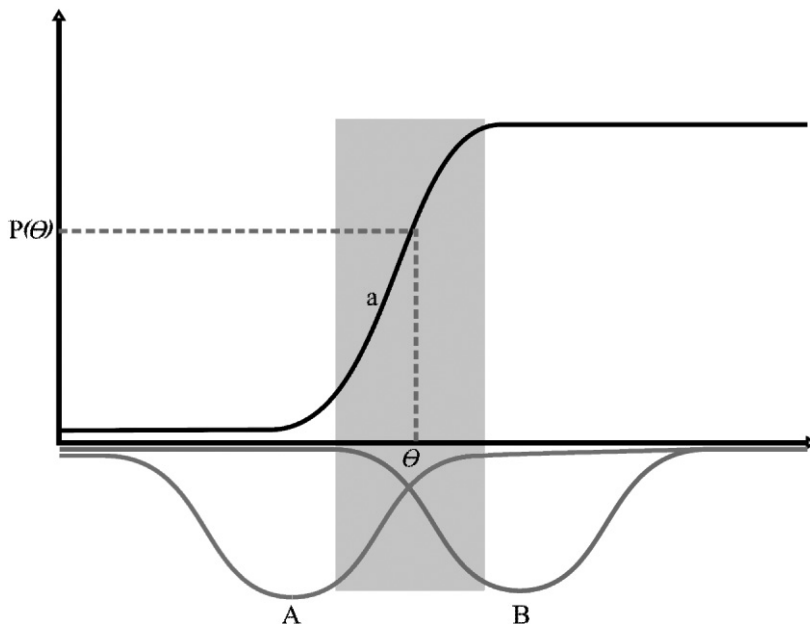


Figura 3. Independência dos Parâmetros dos itens da amostra da população

CURVA CARACTERÍSTICA DO TESTE DO SUJEITO E ERRO DE MEDIDA

Os modelos da TRI permitem também a construção de um gráfico no qual se apresenta a probabilidade de desempenho de um sujeito no teste dado um valor de θ , esse gráfico recebe o nome de Curva Característica do Teste (CCT). Em comparação com as CCI, que são partes centrais na TRI, as CCT's apresentam uma importância menor e sua principal característica é a de funcionar como ligação ou ponto de comparação entre aspectos da teoria clássica e a de resposta ao item. Para sua construção basta que se somem os valores das curvas características dos itens que compõem o teste em questão, a cada nível de θ se somam os valores $P(\theta)$ de cada item para esse nível (Figura 4). Sua expressão matemática é:

$$CCT = \sum_{i=1}^n P_i(\theta), \text{ onde, } n \text{ é o número de itens que compõem o teste.}$$

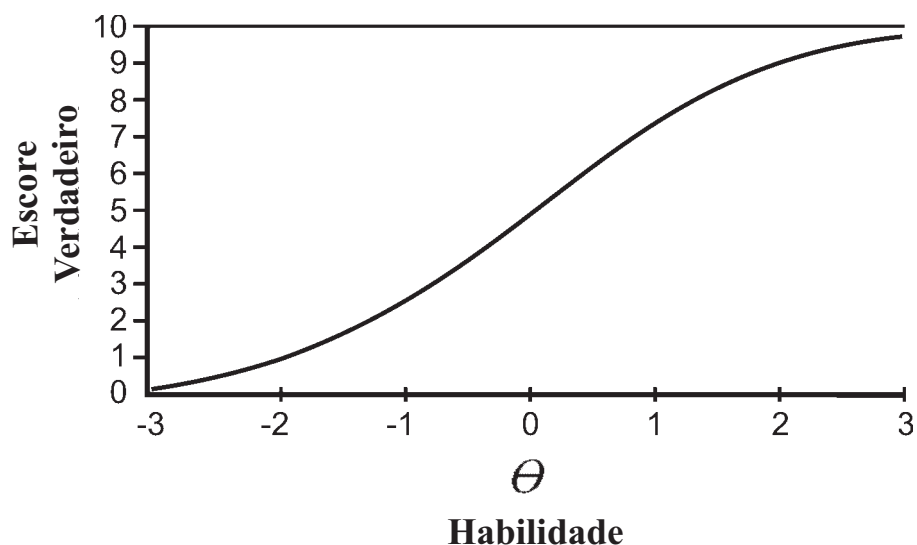


Figura 4. Curva Característica do Teste

O modelo não apresenta como preocupação principal a estimação do escore total do sujeito num determinado teste ou o valor verdadeiro, como é frequentemente chamado na Teoria Clássica. Entretanto, ao se observar a Figura 4, nota-se que no eixo das ordenadas se encontra o equivalente ao valor verdadeiro do sujeito no teste, tal como ele pode ser estimado pela TRI. Esse valor é dado pela soma das probabilidades $P(\theta)$ para cada nível θ presente nas CCI's (Muñiz, 1990).

A utilidade das pontuações verdadeiras ou V estimadas por meio da TRI para a interpretação dos resultados é que elas vêm expressas na mesma escala que as pontuações empíricas, ao passo que os valores de θ são expressos em uma outra escala. O que as CCT's apresentam como pontuações verdadeiras são os valores de θ convertidos para uma escala comum. Na verdade as CCT's demonstram uma relação funcional entre o escore verdadeiro e a escala de habilidade (Baker, 2001).

O cálculo e o uso mais freqüente de θ , em contraposição ao uso do valor do escore verdadeiro, é justamente a independência que se pode obter do valor de θ em relação ao teste utilizado, ao passo que o escore verdadeiro não permite essa independência (Muñiz, 1990).

Da mesma forma que nas CCI's e nas CCT's, podemos a partir dos mesmos dados traçar uma Curva Característica do Sujeito (CCS), que tem como sua principal aplicação a possibilidade de comparação entre os valores empíricos do sujeito com sua curva teórica esperada. Ao serem conhecidos os valores de habilidade dos sujeitos, θ 's, e os parâmetros dos itens que compõem um teste, pode-se estabelecer uma expectativa de desempenho para os sujeitos naquele teste e representá-la graficamente. As CCS's podem ser obtidas representando no eixo das abscissas os valores da dificuldade dos itens divididos em categorias e no eixo das ordenadas, a proporção de acerto do sujeito em itens daquela dificuldade. Então, no mesmo gráfico podemos desenhar a curva com os valores da expectativa de desempenho para o sujeito informando o que ele domina, por exemplo, observando os itens em que ele tem expectativa de mais de 70% de acerto. Outra aplicação para as CCS's seria a possibilidade de comparação de expectativas de desempenho relativas entre vários sujeitos. Por exemplo, a Figura 5 compara duas pessoas em suas capacidades de acertar itens com determinados índices de dificuldade, nota-se que a pessoa A apresenta probabilidade de acerto de 50% nos itens com dificuldade em torno de -1,2, que corresponde a sua faixa de *theta*, ao passo que a pessoa B apresenta probabilidade de acerto de 50% a itens com dificuldade em torno de 1,8, que corresponde a sua faixa de *theta*. Logo se espera que a pessoa B tenha probabilidade de 100% de acerto em itens que a pessoa A apresenta apenas probabilidade de 50% de acerto.

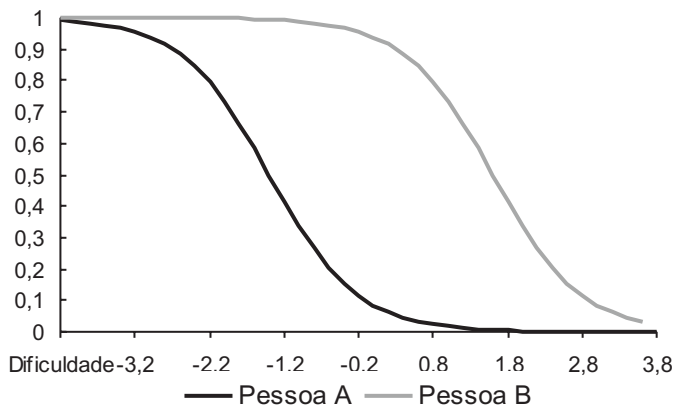


Figura 5. Curvas características de duas pessoas, demonstrando a expectativa de acertos em itens de diferentes níveis de dificuldade

FUNÇÕES DE INFORMAÇÃO

Uma das possibilidades do modelo da TRI diz respeito ao cálculo do erro de medida, que na teoria clássica é dado pela fórmula: $E = T - V$ e o erro padrão de medida é dado pelo desvio pa-

drão das diferenças em relação ao valor verdadeiro, isto é, pelo desvio padrão de E . No modelo de resposta ao item pode-se dizer que o erro também é encontrado utilizando-se recursos analíticos para estimar a variabilidade das estimativas do θ . A característica mais importante sobre a diferença de informação do erro padrão de medida na Teoria Clássica e na TRI é que na segunda, o seu valor não é o mesmo para todos os sujeitos, mas está condicionado ao valor de θ . Isso implica em que a precisão do teste não é a mesma ao longo da escala, pois depende do nível dos sujeitos na variável medida, ou seja, do valor de θ (Baker, 2001).

O Erro Padrão de Medida (EPM) é mais frequentemente estudado nas funções de informação, que são apenas outra forma de se expressá-lo. Portanto a função de informação de um teste é um indicador da precisão desse teste, pois na proporção em que existe mais “informação” sobre o traço medido menor o EPM. A função de informação diz o quão bem cada nível de habilidade pode ser estimado, usando-se um conjunto de itens específico (Baker, 2001).

A expressão gráfica exibe no eixo das abscissas os valores de θ divididos no conjunto de categorias em que foi avaliado e no eixo das ordenadas os valores de $I(\theta)$ (Figura 6). De posse dessa informação pode-se selecionar os testes de melhor capacidade de informação para avaliar sujeitos com relativos valores de θ ou ainda, pensando na construção se dispomos de um conjunto de itens calibrados pode-se construir um teste que apresente uma determinada $I(\theta)$ conforme os objetivos de testagem. Para os testes de medida em geral a configuração do gráfico da função de informação apresentaria uma reta, o que significa uma medida com igual precisão ao longo da escala de habilidade (Baker, 2001).

É possível se produzir um gráfico de função de informação para os itens individualmente ou para o teste como um todo, o que é mais freqüente é o cálculo da função para o teste. Apesar de a teoria de resposta ao item ser uma teoria “itemizada”, ou seja, prioriza a estimação de parâmetros para os itens, o cálculo de uma função de informação para itens individuais raramente é executado (Baker, 2001).

A função de informação dos itens é um poderoso instrumento para análise de itens, possibilitando o conhecimento não só de quanto de informação um item acumula num determinado valor de θ , mas também em que valor de θ o item possui maior quantidade de informação. Segundo Muñiz (1990), a função de informação dos itens tem sido o método de análise de itens mais utilizado pelos construtores de teste atualmente.

Ao analisarmos as funções de informação dos testes e em seguida dos itens podemos deduzir quando elas expressam sua capacidade máxima de informação respectivamente, em relação aos modelos logísticos de um, dois e três parâmetros. Nos modelos de um e dois parâmetros os itens atingem sua capacidade máxima de informação quando θ for igual à dificuldade dos itens ($\theta = b$) e os valores da função de informação são simétricos para cada valor de θ acima ou abaixo do valor de $\theta = b$. No modelo de três parâmetros a quantidade de informação será sempre menor que nos dois outros modelos devido à influência do parâmetro probabilidade de acerto ao acaso c .

CONSIDERAÇÕES FINAIS

Durante as seções anteriores foram expostas as principais características dos modelos para itens dicotômicos da Teoria de Resposta ao Item frequentemente aplicados a testes de habilidades.

Especialmente aquelas que visam responder às deficiências da Teoria Clássica no que respeita ao fato de que as estimativas sobre os sujeitos dependem do teste usado, os parâmetros dos testes dependem da amostra usada na construção e a consideração do escore total como referência de medida. Desta forma foram explanadas as características dos principais modelos, como são estimados os parâmetros dos itens em cada um deles, os principais conceitos concernentes aos respectivos modelos e também os avanços proporcionados pela adoção desses procedimentos. Para um conhecimento mais detalhado dos conceitos, e também outras características não citadas aqui, recomenda-se a leitura de Wright e Stone (2004), Linacre e Wright (2002), Baker (2001) ou Muñiz (1990).

A construção de instrumentos de medida é uma tarefa laboriosa, porém possível e necessária para que o profissional possa ter em mãos instrumentos de qualidade que sirvam como ferramentas complementares ao seu trabalho de avaliação. Desta forma acredita-se que seja desejável aos construtores e usuários de teste um conhecimento, mesmo que elementar das características dos modelos da TRI que permita, se não utilizar o modelo em pesquisas de desenvolvimento de instrumentos, ao menos compreender os conceitos, quando são explanados em trabalhos alheios. O objetivo do presente trabalho foi o de apresentar esses conceitos de forma simples e detalhada o suficiente para uma compreensão e uso práticos, pois o conhecimento dos principais procedimentos de construção, inclusive aqueles oriundos da Psicometria Moderna, ajuda os profissionais a aprimorar a capacidade de realizar julgamentos sobre a qualidade dos instrumentos com base em princípios calcados no método científico. Portanto, ajuda a identificar aqueles que, apesar de prometerem grandes revelações sobre os padrões fixos de comportamento ou dos aspectos subjetivos dos indivíduos, não conseguem comprovar sua utilidade.

REFERÊNCIAS

- Anastasi, A. & Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: Artmed.
- Baker, F.B. (2001). *The basics of item response theory*. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation.
- Embretson, S.E. & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, New Jersey, London: Lawrence Erlbaum Associates.
- García, M.I.B; Maranon, P.P; Falcon, J.C.J & Costas, C.S.L. (2001). Relaciones empíricas entre los estadísticos de la teoría clásica de los testes y los de la teoría de respuesta a los items. *Psicothema*, 13 (2), 324-329.
- Hambleton, R.K. (1990). Item response theory: Introduction and bibliography. *Psicothema*, 2 (1), 97-107.
- Lawley, D.N. (1943). On problems connected with item selection and test construction. *Proceeding of the Royal Society of Edimburg*, 61, 273-287.
- Linacre, J.M. & Wright, B.D. (2002). Understanding Rasch measurement: Construction of measures from many-facet data. *Journal of Applied Measurement*, 3 (4), 486-512.
- Lord, F.M. (1952). A theory of test scores. *Psychometric Monograph*, 7, 17.

- Lord, F.M. (1953). An application of confidence intervals of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 18, 57- 75.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: LEA.
- Muñiz, J. (1990). *Teoría de respuesta a los ítems: Un nuevo enfoque en la evolución psicológica y educativa*. Madrid: Ediciones Pirámide, S. A.
- Pasquali, L. (1996). *Teoria e métodos de medida em ciência do comportamento*. Brasília: INEP.
- Pasquali, L. (1997). *Psicometria: Teoria e aplicações. A teoria clássica dos testes psicológicos*. Brasília: Editora UnB.
- Pasquali, L. (2000). *Psicometria: Teoria dos testes psicológicos*. Brasília: LabPAM.
- Richardson, M.W. (1936). The relationship between difficulty and the differential validity of a test. *Psychometrika*, 1, 33-49.
- Soares, T.M; & Pereira D.R.M. (2002). Estudo de critérios de adequação para modelos da teoria da resposta ao item (TRI) aplicado ao caso do ensino fundamental da micro-região de Juiz de Fora em 1999. *Educação em Foco*, 6 (2), 91-108.
- Tucker, L.R. (1946). Maximum validity of a test with equivalent items. *Psychometrika*, 11, 1-13.
- Wright, B.D. & Stone, M. H. (2004). *Making measures*. Chicago: Phaneron Press.

Recebido em 09/06/09
Revisto em 25/02/11
Aceito em 28/02/11