Psicometria Moderna: características e tendências[1]

WAGNER BANDEIRA ANDRIOLA*

RESUMO

O artigo pretende descrever as novas tendências da Psicometria Moderna, sintetizadas na constatação do uso massivo dos modelos unidimensionais da Teoria da Resposta ao Item (TRI). Para tanto, faz-se um breve relato histórico da TRI, apresentam-se os quatro modelos unidimensionais da TRI, discutem-se os dois principais supostos da TRI – unidimensionalidade e independência local dos itens – e a técnica estatística mais adequada à sua verificação – a análise fatorial. Descrevem-se as fases de estimação dos parâmetros métricos dos itens e da aptidão dos indivíduos, as respectivas provas para verificar o ajuste do modelo aos dados e, finalmente, algumas ideias básicas acerca da função de informação dos itens.
Palavras-chave: psicometria, TRI, avaliação da educação.

RESUMEN

El artículo describe las nuevas tendências de la Psicometría Moderna, sintetizadas en la constatación del uso intensivo de los modelos unidimensionales de la Teoría de la Respuesta al Ítem (TRI). Para ello, es hecho um breve relato histórico de la TRI, son presentados los cuatro modelos unidimensionales de la TRI, son descriptos los dos principales supuestos de la TRI: la unidimensionalidad y la independencia local de los ítems, así como la técnica estadística más adecuada a su verificación (el análisis factorial). Son descriptas las fases de estimación

[1] Trabalho apresentado na I Reunião da Associação Brasileira de Avaliação Educacional (Abave), realizada em Belo Horizonte (maio de 2006).

* Professor Adjunto IV da Universidade Federal do Ceará (UFC); Professor do Programa de Mestrado e Doutorado em Educação; Professor do Mestrado Profissional em Políticas Públicas e Gestão da Educação Superior (POLEDUC/UFC); Coordenador de Avaliação Institucional (SDI/UFC) (w_andriola@yahoo.com/w_andriola@ufc.br).

de los parámetros métricos de los ítems y de la aptitud de los sujetos, las respectivas pruebas para verificar el ajuste del modelo a los datos y, finalmente, algunas ideas básicas a cerca de la función de información de los ítems.
Palabras clave: psicometría, TRI, evaluación de la educación.

ABSTRACT

This paper intends to describe the new trends of Modern Psychometry, summarized in the massive use of the unidimensional models of the Item Response Theory (IRT). We presented a brief historic review of the IRT, the four unidimensional models of IRT, the two main theorical assumptions of the IRT – unidimensionality and local independence of the items – and the statistical technique more adequate to its verification – the factorial analysis. We emphasized the steps for estimation of the metric parameters of items and the individual aptitude, the respective tests for the verification of the adjusted model and, finally, we presented some basics ideas on the item information function.
Keywords: psychometry, IRT, course evaluation.

INTRODUÇÃO

Na área da Física, o conceito de *escala espacial* refere-se ao tamanho de um objeto qualquer, com o qual interagimos num certo momento ou lugar. No nosso cotidiano doméstico lidamos com objetos de diferentes magnitudes: da ordem de alguns milímetros e centímetros (micro-chips, agulhas de costura, pilhas para relógios e máquinas de calcular), de metros (sofás, mesas e camas), de dezenas de metros (divisões de uma casa ou de um apartamento), de centenas de metros (quadras que conformam bairros, que por seu turno constituem cidades). Tais exemplos, propositadamente apresentados, destacam a ideia de que os avanços científicos verificados nos últimos anos possibilitam-nos melhor compreensão da relação existente entre os mundos *microscópio* e *macroscópico*[2]. A esse respeito cabe destacar, como forma de ilustrar a mencionada relação, como o *volume* de um balão – que é uma propriedade associada à escala na qual o balão apresenta sua forma característica – depende das propriedades microscópicas das partículas do gás utilizado para inflar o balão. Conforme asseveram Costa e Bianchi (2002), a extensão das influências ao longo das *escalas espaciais* oferece valioso subsídio para a compreensão e a análise de problemas físicos ou de outras naturezas quaisquer.

A situação descrita é análoga ao que está ocorrendo, atualmente, com uma área da Psicologia e da Educação denominada *Psicometria Moderna*. Ambas as áreas, como qualquer outra ciência, exigem e empregam medições precisas das variáveis que manejam. Assim, igualmente como observamos a tendência da Física em utilizar medidas microscópicas para compreender fenômenos macroscópicos, a Psicometria Moderna vem empregando novos modelos de medida, conhecidos genericamente pelo nome de *Teoria da Resposta ao Item (TRI).*

Como o próprio nome indica, a TRI possui como foco o estudo individualizado dos itens componentes de um grupo – teste ou banco de itens – ao contrário da sua predecessora, a Teoria Clássica dos Testes (TCT), que tinha como objetivo a

---

[2] O prefixo *nano* indica uma unidade de medida derivada, igual a 10^{-9} vezes a primeira (Cf. Dicionário *Novo Aurélio, Século XXI*). No campo da Física, o termo *quark* foi cunhado pelo físico norte-americano Murray Gell-Man e refere-se a um dos constituintes fundamentais da matéria. No seu livro, *El quark y el Jaguar* (4ª edição, Madrid: Ed. Metatemas, 2003), há uma citação que sintetiza muito bem a associação entre o micro e o macro, o simples e o complexo: *(...) los quarks son las partículas elementales que constituyen el núcleo atómico (...). El Jaguar representa la complejidad del mundo que nos rodea, especialmente tal como se manifiesta en los sistemas complejos adaptativos (...). La imagen del quark e del Jaguar transmite perfectamente mi idea de lo simple y lo complejo: de un lado las leyes físicas subyacentes de la materia y el universo, y del otro, el rico entramado del mundo que percibimos directamente y del que formamos parte* (p. 29).

determinação das propriedades ou dos parâmetros métricos do teste (Muñiz, 1997). Podemos afirmar, nesse âmbito, que a TRI preocupa-se com o estudo das características métricas dos itens, utilizando, para tanto, uma *escala microscópica*; já a TCT tem seu foco direcionado ao próprio instrumento de medida – o teste – e emprega, para tal, uma *escala macroscópica*. É exatamente nesse aspecto em que reside a mais significativa distinção entre a TRI e a TCT (Hambleton, 1997; Andriola, 2002).

Sendo um novo modelo teórico, necessita que muitos dos conceitos e dos supostos sejam melhor explicados pelos psicometristas. Nosso texto tem a pretensão de lançar mais luzes sobre essa área, ainda muito superficialmente conhecida na Psicologia e na Educação, sobretudo no Brasil. Assim, realizaremos breve descrição da TRI, iniciando com comentários superficiais acerca dos principais colaboradores e dos responsáveis pelo seu surgimento, pois conforme opinou o matemático e filósofo britânico *Sir* Alfred North Whitehead (1861-1947): *uma ciência que esquece dos seus fundadores está irremediavelmente perdida*.

TEORIA DA RESPOSTA AO ITEM (TRI): BREVE RETROSPECTIVA HISTÓRICA

Como sempre ocorre no âmbito das Ciências, os modelos ou teorias não surgem de um momento para outro, da noite para o dia, nem tampouco de forma linear (Goldstein; Wood, 1989; Serres, 1998). Conforme destaca Muñiz (1994), a TRI não é, como se poderia ingenuamente acreditar, um novo enfoque psicométrico, ainda que alcance solucionar certos problemas da TCT. Lord (1980) vai um pouco mais além e afirma que a TRI não contradiz nem os supostos nem as conclusões fundamentais da TCT, tão somente apresenta supostos adicionais que permitem responder questões que a TCT não podia responder.

Nesse contexto, desejamos ressaltar que o processo de criação e posterior evolução da TRI foi lento, difícil e sinuoso, tendo começado há mais de 50 anos (Hambleton, 1990; Muñiz; Hambleton, 1992; Van der Linden; Hambleton, 1997). O artigo de L. L. Thurstone, publicado em 1925, no qual é apresentada uma série de curvas, associando a idade dos indivíduos com a proporção de acertos de cada item, é considerado o antecedente das modernas *curvas características dos itens (CCI's).* Posteriormente, M. W. Richardson, em 1936, tentou ajustar a ogiva normal às respostas aos itens, ademais de preocupar-se em adaptar a dificuldade dos itens aos objetivos do teste, o que representa a formulação antecipada da *função de informação do item (FI).*

G. A. Fergunson, em 1942, também se aproximou do conceito de CCI, utilizando os métodos psicofísicos, mais precisamente o *método dos estímulos constantes*. Suas investigações possibilitaram o desenvolvimento de definições equivalentes aos do *parâmetro de dificuldade* (*parâmetro b*), isto é, o *valor da variável latente medida* (geralmente simbolizada pela letra grega θ) quando *a probabilidade de acertá-la é idêntica à de errá-la* [em notação matemática: $P(\theta) = 0{,}5$], supondo-se a inexistência de acertos ao acaso. Já em 1950, P. F. Lazarsfeld utilizou, pela primeira vez, o termo *traço latente*, nome que seria adotado para designar os modelos da TRI. Em síntese, estas são as origens mais remotas do nascimento formal da TRI.

Cabe destacar, no entanto, o trabalho de F. M. Lord, de 1952, resultado de sua tese doutoral dirigida por H. Guliksen e assessorada por L. R. Tucker, como a formulação mais sistemática dos principais conceitos e modelos da TRI. Entre os anos 1957-58, A. Birnbaum propôs novos modelos logísticos, fundamentados nos da ogiva normal de F. M. Lord, cuja vantagem mais visível é a maior facilidade de tratamento matemático que, por conseguinte, possibilitou a geração de novos procedimentos visando a sua aplicação prática (Van der Linden; Hambleton, 1997).

O dinamarquês G. Rasch publicou em 1960 um trabalho no qual expôs com ricos detalhes o modelo logístico de um parâmetro. O mencionado autor introduziu dois novos princípios à moderna teoria da medida, a TRI: *convergência* e *separabilidade*. Conforme palavras de Gavíria Soto (2000):

> El primero de ellos tiene que ver con la idea de que es imposible resolver el problema de la medida en las ciencias sociales si no se entra en el círculo no vicioso, en la circularidad hermenéutica de una formalización simultánea o conjunta de los parámetros de las preguntas y sus respuestas. (p. 18)

Assim, o *princípio de convergência* fundamenta-se na ideia de que, sobre as bases de uma primeira equação do modelo de medida, podem ser estimados os parâmetros dos itens independentemente da magnitude das pessoas na variável latente (θ), tendo sido trocado este último aspecto por algo "diretamente observável": o número total de respostas corretas (os escores obtidos no grupo de itens empregado). Posteriormente, sobre as bases de uma segunda equação, pode-se estimar a magnitude das pessoas na variável latente (θ) conhecendo-se apenas os parâmetros dos itens, que foram estimados na primeira equação. Finalmente, um terceiro modelo permite o contraste dos resultados da primeira com os da segunda equação. Esta terceira equação é independente de todos os parâmetros (dos itens e da magnitude das pessoas na variável latente θ), sendo dependente, tão somente, do número total de respostas corretas.

Já o segundo princípio, o da *separabilidade*, está associado à possibilidade de comparar as habilidades entre indivíduos, sem necessidade de fazer referência ao instrumento de medida utilizado (Gavíria Soto, 2000). No modelo de G. Rasch estão presentes dois supostos básicos: 1º) o indivíduo com maior habilidade deveria ter, igualmente, maior probabilidade de acertar qualquer item ou problema; 2º) o item mais fácil deveria possuir, teoricamente, maior probabilidade de ser respondido corretamente.

Mais adiante, em 1968, F. M. Lord e M. R. Novick incrustam, definitivamente, os seus nomes na história da psicometria moderna. Ambos publicam o livro *Statistical Theories of Mental Tests Scores*, marcando, desse modo, o final da primeira fase dos modelos TRI que ainda estavam no nível teórico-matemático. Em razão de sua complexidade matemática, da ausência de *softwares* específicos para o seu uso e, sobretudo, pela descrença generalizada acerca de suas reais vantagens, os modelos TRI teriam que esperar duas longas décadas para serem mundialmente aceitos e empregados, conforme atesta Hambleton (1990).

Em 1980 outra publicação do gênio F. M. Lord – *Applications of Item Response Theory to Practical Testing* – lançou as bases para o uso da TRI nos campos da avaliação psicológica e educacional. A partir de então, ocorreu verdadeira expansão e fortalecimento da TRI, com o reconhecido predomínio de seu uso na psicometria moderna. Concomitantemente, multiplicaram-se as publicações científicas[3] e as revistas especializadas na área, os congressos abordando a avaliação por meio da TRI, dentre outras formas de abordagem da temática. Para concluir, cabe uma constatação: *na atualidade, a psicometria possui um enfoque teórico dominante, o da TRI.*

PRINCIPAIS MODELOS LOGÍSTICOS DA TRI

O modelo mais parcimonioso foi proposto por G. Rasch, em 1960, tendo recebido o nome de *modelo logístico de um parâmetro* (Andriola; Barreto, 1997). De acordo com esse modelo, a probabilidade de acerto a um item é influenciada pela sua dificuldade (*parâmetro b*). Expressando-o em notação matemática temos: $P(\theta) = \dfrac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$. Nela:

- *P(θ)* é a probabilidade de acerto do item *i,* dada determinada magnitude de θ;
- *θ* é a variável latente medida pelo item *i*;

3. Como ilustração do amplo interesse pela TRI, mencione-se que a partir de 1980 mais de 300 artigos foram publicados por dois dos mais prestigiados periódicos internacionais: *Journal of Educational Measurement* e *Applied Psychological Measurement*.

- b_i é o índice de dificuldade do item *i*;
- *e* é o valor (2,72) correspondente à base dos logaritmos neperianos;
- *D* é uma constante de valor 1,7.

O parâmetro *b* do modelo em questão corresponde ao valor de θ no ponto de máxima inclinação da CCI, estando o seu valor numérico na mesma escala de medida de θ. Neste caso, mantendo-se as condições de normalidade de θ, isto é: *N(0, 1)*, o parâmetro *b* relaciona-se com o índice de dificuldade da TCT, sendo dado pela expressão matemática: $b \cong \frac{-Z_p}{r_b}$, onde Z_p é o escore padronizado, que corresponde, na curva normal, à proporção de acerto ao item em foco e que é denominado, na TCT, de índice de dificuldade do item. Já o termo r_b é a correlação bisserial entre a pontuação obtida no item e o escore total obtido no teste.

O segundo tipo é denominado *modelo logístico de dois parâmetros* e foi proposto por A. Birnbaum, tendo sido desenvolvido entre 1957 e 1968. Assume que a probabilidade de acerto a um item é influenciada pela sua dificuldade (*parâmetro b*) e pela sua discriminação (*parâmetro a*). Em termos matemáticos, o modelo é expresso por $P(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$, onde a_i é o *índice de discriminação do item i* e os demais símbolos [$P(\theta)$, θ, *e*, b_i, *D*] assumem o mesmo significado do modelo anterior, proposto por G. Rasch.

O valor do *parâmetro a* é proporcional à inclinação da reta tangente à CCI, no ponto de maior valor desta. Em outras palavras: quanto maior for a inclinação da curva, maior será o valor do *parâmetro a* e, portanto, maior o seu poder para diferenciar sujeitos que possuam distintas magnitudes de θ. Ademais, quando θ tem distribuição normal [*N(0, 1)*] e não há acertos ao acaso (isto é: quando c = 0), o valor do *parâmetro a* é dado, aproximadamente, pela expressão:

$a \cong \frac{r_b}{\sqrt{1-(r_b)^2}}$, onde r_b é o valor da correlação bisserial entre o item e o escore total, ou seja, o o equivalente ao *índice de discriminação*, no âmbito da TCT.

O *modelo logístico de três parâmetros* foi também desenvolvido a partir dos trabalhos pioneiros de A. Birnbaum e assume que a probabilidade de acerto a um item é influenciada pela sua dificuldade (*parâmetro b*), seu poder de discriminação (*parâmetro a*) e pela chance, em termos probabilísticos, de que seja acertado ao acaso (*parâmetro c*). Matematicamente, o modelo é expresso por:

$P(\theta) = c_i + (1 - c_i)\dfrac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$, na qual c_i é *a probabilidade de que o item i seja acertado ao acaso*. Os demais símbolos [$P(\theta)$, θ, e, b_i, D] assumem o mesmo significado dos dois modelos anteriores.

Neste modelo o *parâmetro c* representa o valor assintótico da CCI quando θ tende a -∞. Em palavras menos técnicas, o mencionado valor representa a probabilidade de acertar o item ao acaso, isto é, quando o respondente nada sabe acerca do que está sendo avaliado pela questão ou item; quando a magnitude do sujeito na variável latente θ medida pelo item é muito pequena, isto é, quando tende a -∞. Seu equivalente na TCT recebe a mesma denominação, isto é, *probabilidade de acerto ao acaso,* porém com distinto modo de determinação matemática. Por exemplo, suponhamos um item que tenha cinco alternativas propostas como respostas, sendo tão somente uma delas a correta. Na TCT a probabilidade de acerto ao acaso viria dada por $p = \dfrac{AC}{AP}$, onde *AC é o número de alternativas corretas* e *AP é o número de alternativas propostas*. Assim, teríamos p = 1 / 5 = 0,2.

Finalmente, o modelo logístico de quatro parâmetros foi proposto por M. A. Barton e F. M. Lord, em 1981 para investigar o problema de que, algumas vezes, por circunstâncias muito específicas, tais como o descuido do respondente em escolher a alternativa correta ou o uso de procedimentos muito específicos por parte do elaborador do item, os sujeitos com elevada competência (ou elevada magnitude na variável latente medida pelo item, o θ) não conseguem acertá-lo. Por isso, os autores incorporaram o parâmetro Y_i, sendo sua formulação matemática a seguinte:

$P(\theta) = c_i + (Y_i - c_i)\dfrac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$, no qual o termo Y_i adota valores inferiores a 1 e os demais componentes da equação são os mesmos descritos para os modelos de um, dois e três parâmetros logísticos.

Atualmente, são poucas as investigações que utilizam esse modelo. A principal justificativa para isso, segundo Muñiz (1997), reside no fato de que o mesmo não apresenta vantagens significativas comparativamente aos outros três modelos e, ademais, os problemas que trata de solucionar podem ser muito bem controlados durante a elaboração dos itens. Descritos os quatro modelos logísticos, passemos agora a descrever os dois supostos centrais da TRI: a *unidimensionalidade* e a *independência local dos itens*.

UNIDIMENSIONALIDADE E INDEPENDÊNCIA LOCAL DOS ITENS

Ao principal suposto da TRI, conhecido como unidimensionalidade, subjaz uma ideia simples e atraente, que podemos encontrar em antigas concepções gregas, conforme assevera Dunham (2002):

> A los antiguos griegos les apasionaban las simetrías, la belleza visual y la sutil estructura lógica de la geometría. Especialmente fascinante les resultaba en como lo simple y elemental podían servir de fundamento a lo complejo e intrincado. (p. 34)

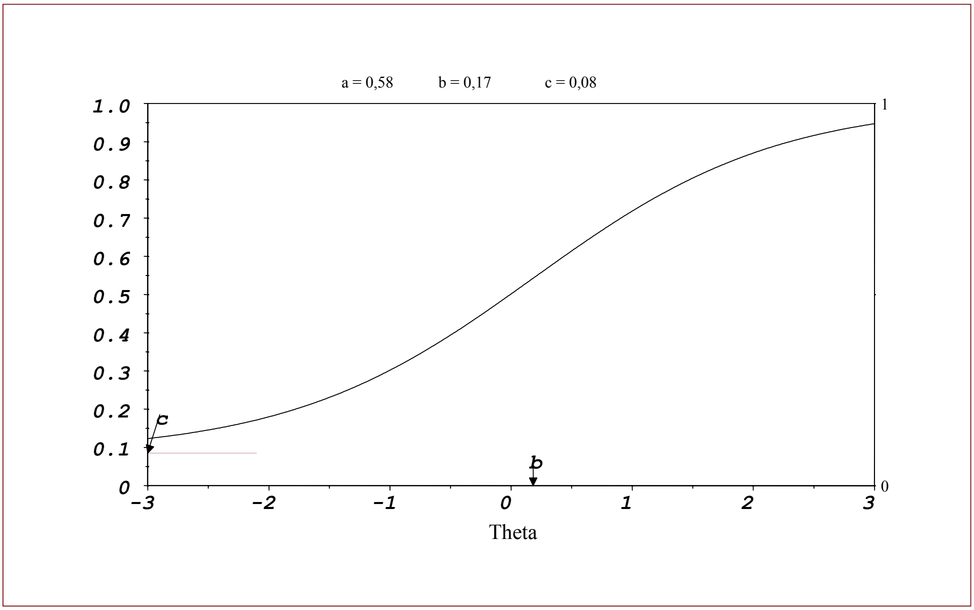
A unidimensionalidade é uma proposição teórica parcimoniosa e elegante, segundo a qual toda a complexidade intrínseca ao ato de resolução de um problema – de natureza cognitiva ou não – deve ter como causa uma única estrutura latente, denominada θ. Assim, *existirá uma relação funcional entre θ e os padrões das respostas dadas a um problema*, sendo essa a ideia fundamental dos quatro modelos logísticos de TRI já descritos, conforme destacam Hambleton (1983, 1990), Van der Linden e Hambleton (1997). Desse modo, a probabilidade de um indivíduo *j* acertar ao item *i* dependerá, exclusivamente, da magnitude que ele possua na variável latente medida θ e das características do item (β_i) que sejam consideradas pelo modelo – dificuldade, discriminação ou acerto ao acaso.

A relação funcional referida pode vir a ser representada graficamente pela *curva característica do item (CCI)*. Como destaca Pasquali (1997), a CCI é a característica particular de cada item, é sua carteira de identidade, uma vez que compartilha uma forma geral muito parecida com um **S**. A figura 1 apresenta a CCI de um item cujos parâmetros foram estimados pelo modelo logístico de três parâmetros.

Conforme propõem Roussos, Schnipke e Pashley (1999), a probabilidade de acerto é expressa por: $P\ [X_{ij} = 1 \mid \theta_j, \beta_i = P_i\ (\theta_j)]$. Já a probabilidade de errar a resposta será dada por: $P\ [Q_i\ (\theta_j) = 1 - P_i\ (\theta_j)]$. Desse modo, espera-se que sujeitos com distintas magnitudes na variável latente medida tenham diferentes probabilidades de acertar a um mesmo item, conforme representado na figura 1.

Ressaltada a relação funcional teorizada pelos modelos unidimensionais da TRI, passaremos a analisar seus dois principais supostos teóricos. Comecemos afirmando que a *independência local dos itens* baseia-se na ideia de que a resposta a um item qualquer não afeta as respostas posteriores fornecidas aos demais itens de um grupo maior. Em palavras menos técnicas: *os itens de um teste não podem apresentar pistas que permitam aos respondentes acertar outros itens, posteriormente apresentados*. De acordo com essa ideia, a independência local dos itens pode ser matematicamente

Figura 1 – Curva característica do item (CCI) e seus respectivos valores para os parâmetros de discriminação *(a)*, dificuldade *(b)* e acerto ao acaso *(c)*



definida como o produto das probabilidades de acertar a cada um dos itens que compõem um teste unidimensional, isto é: $P(U_1, U_2, ..., U_n \mid \theta) = P(U_1|\theta) \times P(U_2|\theta) \times ... \times P(U_n|\theta)$, na qual U é a probabilidade de acerto a um determinado item, dado certa magnitude de θ.

Cabe destacar um paradoxo interno a este suposto teórico, caso houvesse o descumprimento da independência local dos itens. *Ocorreria que a variável unidimensional θ não teria toda a sua variância explicada pelos itens utilizados para medi-la*. Desse modo, far-se-ia necessário empregar outros itens que medissem aspectos não contemplados pelos itens já utilizados. Como resultado, conforme assevera Harrison (1986), a variância a ser explicada já não dependeria somente dos itens inicialmente utilizados, senão também de um conjunto secundário de itens. Nesse caso, a unidimensionalidade da variável latente estará ameaçada.

Dessa contradição, é possível inferir que, sendo respeitado o suposto da unidimensionalidade, pode-se corroborar, matematicamente, a independência local dos itens. Assim, comprovamos a equivalência entre os conceitos de unidimensionalidade e independência local dos itens, conforme a opinião de Goldstein (1980). O problema agora reside numa indagação de cunho técnico: *qual o procedimento estatístico mais adequado à verificação da unidimensionalidade de um grupo de itens?*

IDEIAS BÁSICAS ACERCA DA ANÁLISE FATORIAL

Na opinião de García Jiménez, Gil Flores e Gómez (2000), determinar a estrutura fatorial que subjaz a um grupo de itens é algo extremamente importante para uma atividade com pretensões científicas. Conforme Keeling (2000):

> Determining the dimensionality of data from an empirical study is crucial to the interpretation of the analyses. Researchers can select from various rules to determine the 'correct' number of dimensions in a data set. (p. 457)

De acordo com Gavíria Soto (1988), Hattie (1984; 1985), Hattie e outros (1996), Keeling (2000) e Nandakumar (1994), embora exista enorme diversidade de métodos para a determinação da estrutura fatorial de um conjunto de dados – *dentre os quais podem ser destacados: o procedimento de Bejar; o contraste de Gustaffson; o método de McDonald; o contraste* Q_1 *e* Q_2 *de Van den Wollenberg; a análise de precedência modificada; o método Hattie para a comparação de autovalores reais e simulados; o método da equação de regressão* – a maioria dos autores ainda recomenda o emprego da análise fatorial, cuja formulação matemática é:

$X_i = a_iF_1 + a_iF_2 + \ldots + a_kF_k + u_iD_i$, onde:

- X_i é a pontuação obtida na variável observada i;
- a_iF_1 é a carga fatorial da variável observada X_i no fator 1;
- a_iF_2 é a carga fatorial da variável observada X_i no fator 2;
- a_kF_k é a carga fatorial da variável observada X_i no fator k;
- u_iD_i é a unicidade (variância específica ou não compartilhada) da variável observada X_i no fator específico D_i.

De acordo com Martínez Arias (1997), esse modelo linear tenta explicar, assim, a atuação de construtos latentes ou fatores (F) sobre variáveis observadas (X), cada uma delas com seus pesos específicos, intensidades, saturações ou cargas fatoriais (a). Como é óbvio, a totalidade da variância nunca é explicada pelos fatores comuns (F), pois existe uma parte (variância não explicada) que se deve aos fatores específicos (D), que também exercem influência com certa intensidade (u) sobre as variáveis observadas (X).

Os modelos de TRI unidimensionais têm como objetivo extrair tão-somente um fator para sintetizar o modelo geral $X_i = a_iF_1 + u_iD_i$. O problema é que raras vezes encontra-se unidimensionalidade perfeita, ou seja, tão somente um fator que explique 100% da variância total. Desse modo, a unidimensionalidade converte-se numa questão de graduação, isto é, quanto mais variância explique o primeiro fator, maior grau de unidimensionalidade existirá, conforme assevera Andriola (2002).

PRINCIPAIS DIFICULDADES NO USO DA ANÁLISE FATORIAL

Abundantes estudos têm sido executados para avaliar a robustez dos modelos de TRI quando estes violam o suposto da unidimensionalidade. A maioria desses estudos utiliza dados simulados por computador, conforme asseguram Ansley e Forsyth (1985), Drasgow e Parsons (1983) e Reckase (1979). Geralmente, como era de se esperar, o problema do não cumprimento do citado suposto diminui à medida que aumenta a variância explicada pelo primeiro fator. Trabalhos realizados por autores como Cuesta (1996), Drasgow e Parsons (1983) indicam que os modelos unidimensionais de TRI são bastante resistentes às violações do mencionado suposto teórico. Não obstante, têm sido desenvolvidos vários procedimentos estatísticos tendo por objetivo melhorar as estimações com base no uso da matriz de correlações a ser submetida à análise fatorial, todos eles considerando a natureza dos dados, isto é, o seu nível de medida (Christoffersson, 1975; Hair Jr. et al. 1999; Keeling, 2000; Stout et al. 1992).

Autores como Bernstein e Teng (1989), McDonald e Ahlawat (1974), Mislevy (1986) e Muthén (1978, 1989) opinam que o modelo linear de análise fatorial produz resultados enviesados quando é utilizado com dados politômicos ou categóricos. Não obstante, muitos investigadores sociais utilizam o *coeficiente* Φ (coeficiente de Pearson para dados dicotômicos) para gerar a matriz de correlação a ser fatorizada. De acordo com Waller (1995), é necessário reconhecer que o modelo linear de análise fatorial pode distorcer a estrutura subjacente ou latente aos dados dicotômicos por duas razões:

- o Coeficiente de correlação de Pearson para dados de natureza dicotômica (coeficiente Φ) é influenciado pela forma de distribuição dos itens, bem como pelo seu conteúdo;
- a dimensionalidade da matriz de correlações de coeficientes Φ pode diferir da verdadeira dimensionalidade da variável subjacente ou latente aos dados. Por exemplo, McDonald (1981), McDonald e Ahlawat (1974) demonstraram que o modelo linear de análise fatorial pode produzir fatores espúrios quando não há uniformidade na dificuldade dos itens. Gavíria Soto (1988) e McDonald (1971, 1999) também destacam que a análise fatorial foi proposta para ser utilizada na análise de variáveis intervalares; quando é utilizada com dados dicotômicos os fatores extraídos são denominados *fatores de dificuldades dos itens.*

De acordo com Wherry e Gaylord (1944) existem duas causas responsáveis pela aparição dos mencionados *fatores de dificuldades*:

- ambiguidade na definição do que se entende por *homogeneidade,* refere-se ao conteúdo, à dificuldade ou a ambas;
- falha na determinação do coeficiente de correlação utilizado para gerar a matriz de intercorrelações.

Tais autores propõem dois procedimentos para resolver essas dificuldades:

- se os itens são homogêneos, tanto no conteúdo como na dificuldade, não importa qual o coeficiente de correlação a ser utilizado, pois não aparecerão fatores espúrios;
- se os itens são homogêneos no conteúdo, mas não na dificuldade, só pode ser utilizada a correlação tetracórica, pois de outro modo aparecem fatores de dificuldade. Como afirmam Knol e Berger (1991), a análise fatorial de dados dicotômicos deve partir sempre da análise de uma matriz de coeficientes de correlações tetracóricas. Gavíria Soto (1988), Lord e Novick (1968) apresentam excelente explicação matemática acerca desse tema.

Nesse âmbito, vale a pena mencionarmos que o programa estatístico *MicroFACT,* desenvolvido por Waller (1995), é bastante simples e extremamente aconselhado na verificação da unidimensionalidade de dados dicotômicos. O referido programa organiza uma matriz secundária a ser fatorizada com base na matriz original de correlações tetracóricas, denominada *smoothed tetrachoric correlation matrix.* Resolvido o problema da geração da matriz de correlações a ser fatorizada, surge uma nova dificuldade: *qual o número de fatores que devemos reter?*

DETERMINAÇÃO DO NÚMERO DE FATORES A RETER

Acerca dessa problemática, Gavíria Soto (1988) expressou sua opinião nos seguintes termos:

> El problema de evaluar la unidimensionalidad de un conjunto de ítems por medio del análisis factorial no es más que un caso particular de un problema ampliamente tratado en la literatura sobre el tema. Se trata de la determinación del número de factores a retener en una solución factorial. (p. 226)

A dificuldade reside, sobretudo, na rotação dos fatores, pois a retenção de mais ou menos fatores afeta a comunalidade das variáveis que, por sua vez, influencia a caracterização dos fatores extraídos, especialmente quando se executam rotações oblíquas. Existe enorme variedade de métodos para determinar o número de fatores a reter depois da fatorização, porém, em razão de seu uso frequente e a ampla citação na literatura mundial, descreveremos os dois mais conhecidos (Andriola, 2002; García Jiménez; Gil Flores; Gómez, 2000).

MÉTODO DE KAISER-GUTTMAN

O método está baseado nas ideias de L. Guttman discutidas no artigo *"Some necessary conditions for common factor analisys"*, publicado na revista *Psychometrika*, em 1954. Posteriormente, H. F. Kaiser publicou, em 1961, o texto *"A note on Guttman's lower bound for the number of commnon factors"*, na revista *British Journal of Statistical Psychology,* que resultou num progresso das ideias iniciais de L. Guttman. O critério consiste em reter aqueles fatores com autovalores (*eigenvalues*) maiores que um, e está fundamentado na ideia de que se um fator é comum deve conter ao menos a variância equivalente a uma variável. No entanto, há grandes discrepâncias acerca da adoção desse critério, uma vez que alguns autores consideram que o seu uso subestima o número de fatores, enquanto outros creem que o superestima (Gavíria Soto, 1988; Keeling, 2000).

MÉTODO *SCREE-PLOT*

Foi proposto por R. B. Catell, cujas ideias foram publicadas em 1966 sob o título *"The scree tests for the numbers of factors"* na revista *Multivariate Behavioral Research.* Trata-se de um procedimento muito simples, fundamentalmente gráfico (Keeling, 2000). Como menciona Gavíria Soto (1988), uma vez obtidos os autovalores para os fatores, estes são situados em um sistema cartesiano, no qual a abscissa representa os sucessivos fatores e a ordenada a magnitude de cada autovalor. O procedimento consiste em traçar uma reta paralela aos fatores que possuam autovalores mais baixos, até que a mesma *"corte"* o eixo das ordenadas. São retidos tantos fatores quanto o número de *eigenvalues* (autovalores) que esteja na parte superior da reta. Não obstante, esse procedimento apresenta problemas quando as diferenças entre as magnitudes dos autovalores correspondentes aos fatores comuns e os fatores únicos são muito pequenas.

Chegado a esse ponto, no qual os fatores foram extraídos, caberá, então, ao investigador estimar os parâmetros dos itens e a aptidão dos indivíduos na variável latente medida (θ).

ESTIMAÇÃO DOS PARÂMETROS DOS ITENS E DA APTIDÃO DOS INDIVÍDUOS NA VARIÁVEL LATENTE

Mencionamos que paralelamente à estimação dos parâmetros dos itens, segundo o modelo TRI escolhido pelo investigador, temos que calcular o valor da variável latente medida (θ) para cada sujeito. A lógica deste procedimento é encontrar como valores para os parâmetros aqueles que maximizem a probabilidade de ocorrência das respostas dos sujeitos aos itens. Conforme assevera Lord (1986), tal método denomina-se *Máxima Verossimilhança,* pelo fato de que os valores estimados são aqueles que tornam mais verossímeis ou plausíveis os dados obtidos empiricamente – as respostas dos indivíduos a cada item. Nesse procedimento, a estimação é feita por aproximações sucessivas denominadas *iterações,* cujo cálculo é muito laborioso e tedioso, pelo que é necessário usar programas estatísticos tais como o *BILOG for Windows*.

As iterações se detêm quando os valores estimados para os parâmetros convergem, isto é, quando após a iteração *n* não se produz mudanças significativas sobre as estimações. O problema é que desconhecemos os valores da variável latente (θ) de cada indivíduo, além dos próprios valores dos parâmetros dos itens, por isso devemos estimá-los ao mesmo tempo, por meio do processo conhecido como *estimação conjunta de máxima verossimilhança*. Os parâmetros a estimar no modelo logístico de três parâmetros será $3n + N$, onde n é o número de parâmetros do modelo e N é o número de sujeitos, dos quais teremos que determinar seus respectivos valores na variável latente (θ). Como já mencionamos, o procedimento adotado nessa estimação consiste em obter conjuntamente os valores dos parâmetros que maximizem a seguinte função matemática:

$$L(u \mid \theta, a, b, c) = \prod_{a=1}^{N} \prod_{i=1}^{n} [P_{ia}(\theta)]^{u_{ia}} [Q_{ia}(\theta)]^{(1-u_{ia})}$$

Para encontrarmos os valores de θ, a, b e c que maximizem a função L temos que resolver a equação $\frac{\partial \ln L}{\partial P_k} = 0$ na qual P é o vetor de parâmetros a estimar P' = [θ, a, b, c] e k é o número de vetores. Hambleton, Swaminathan e Rogers (1991) apresentam as respectivas expressões para a função L, no caso de utilizar-se o modelo logístico de três parâmetros:

Parâmetro a: $$\frac{\partial \ln L}{\partial a_i} = \left[\frac{D}{(1-c_i)}\right] \sum_{a=1}^{N} \left[\frac{(\theta_a - b_i)(P_{ia}(\theta) - c_i)(u_{ia} - P_{ia}(\theta))}{P_{ia}(\theta)}\right]$$

Parâmetro b: $$\frac{\partial \ln L}{\partial b_i} = \left[\frac{D_{ai}}{(1-c_i)}\right] \sum_{a=1}^{N} \left[\frac{(P_{ia}(\theta)-c_i)(u_{ia}-P_{ia}(\theta))}{P_{ia}(\theta)}\right]$$

Parâmetro c: $$\frac{\partial \ln L}{\partial c_i} = \left[\frac{1}{(1-c_i)}\right] \sum_{a=1}^{N} \left[\frac{\dfrac{u_{ia}-(P_{ia}(\theta))}{P_{ia}(\theta)}}{P_{ia}(\theta)}\right]$$

Construto θ: $$\frac{\partial \ln L}{\partial \theta_a} = D \sum_{i=1}^{N} \left[\frac{a_i(P_{ia}(\theta)-c_i)(u_{ia}-P_{ia}(\theta))}{(1-c_i)(P_{ia}(\theta))}\right]$$

Afirmamos que os quatro parâmetros (θ, a, b, c) devem ser estimados conjuntamente. Por isso, será necessário empregar-se tratamento multivariado do procedimento de Newton-Raphson, conforme apregoa Lord (1986). Nesse caso, a estimação é realizada em dois momentos distintos e hierárquicos:

- *Fase 1:* estimação das pontuações θ dos sujeitos;
- *Fase 2:* conhecidas as pontuações θ dos sujeitos são estimados os parâmetros métricos dos itens, com base no modelo logístico que melhor se ajuste aos dados empíricos.

Ambas as fases se repetem até a obtenção da convergência conjunta, e os valores com os quais, finalmente, se alcança a mencionada convergência são, então, adotados como os estimadores de máxima verossimilhança conjunta. Terminada tal etapa, caberá ao investigador testar o ajuste do modelo escolhido aos dados empíricos.

PROVAS PARA VERIFICAR O AJUSTE DO MODELO AOS DADOS EMPÍRICOS

Os principais procedimentos ou provas estatísticas habitualmente utilizados para verificar o ajuste do modelo aos dados empíricos são: o qui-quadrado e a análise de resíduos padronizados. A lógica do qui-quadrado consiste na comparação dos valores prognosticados ou teorizados pelo modelo com os obtidos empiricamente. Para isso, divide-se a variável medida em categorias e comparam-se os valores prognosticados com os empíricos, dentro de cada categoria. No caso do ajuste perfeito, ambos os valores coincidirão em todas as categorias, do contrário não houve ajuste. O que indica o teste do qui-quadrado é se essas diferenças são estatisticamente significativas, conforme expressa Andriola (2002). Sua formulação matemática é dada por $Q_1 = \sum_{j=1}^{k} \frac{n_j [P(\theta_j) - P_e(\theta_j)]^2}{P(\theta_j)[1 - P(\theta_j)]}$ na qual:

- Q_1 tem distribuição como χ^2 com $k - p$ graus de liberdade;
- k é o número de categorias na qual θ foi dividida;
- p é o número de parâmetros do modelo TRI utilizado;
- n_j é o número de sujeitos dentro de cada j categoria;
- $P_e(\theta_j)$ é a proporção de sujeitos que acertou o item, dentro de uma determinada categoria j;
- $P(\theta_j)$ é a probabilidade de acertar o item, dado θ_j.

O segundo modo de verificar o ajuste do modelo aos dados empíricos pode ser pela análise dos resíduos. Também, nesse caso, divide-se θ em várias categorias ou níveis, calculando-se, em seguida, o seu respectivo resíduo. Sua formulação matemática é dada por $RE = \frac{P(\theta_j) - P_e(\theta_j)}{\frac{\sqrt{P(\theta_j)Q(\theta_j)}}{n_j}}$, na qual:

- N_j é o número de sujeitos na categoria j;
- $P(\theta_j)$ é o valor da CCI para o nível θ_j;
- $P_e(\theta_j)$ é proporção de sujeitos que acertou o item na categoria θ_j;
- $Q(\theta_j)$ é o valor resultante de $1 - P(\theta_j)$.

A interpretação dos valores RE indica que à medida que se distanciam de zero, em valor absoluto, pior será o ajuste do modelo aos dados empíricos. Uma inspeção do tamanho dos resíduos para as distintas categorias nas quais se dividiu a variável latente *(θ)* pode dar-nos ideia das zonas de maior ajuste ou desajuste. É frequente o estabelecimento de intervalos de valores admissíveis para RE, por exemplo, RE ≤ 2.

FUNÇÃO DE INFORMAÇÃO DO ITEM (FI)

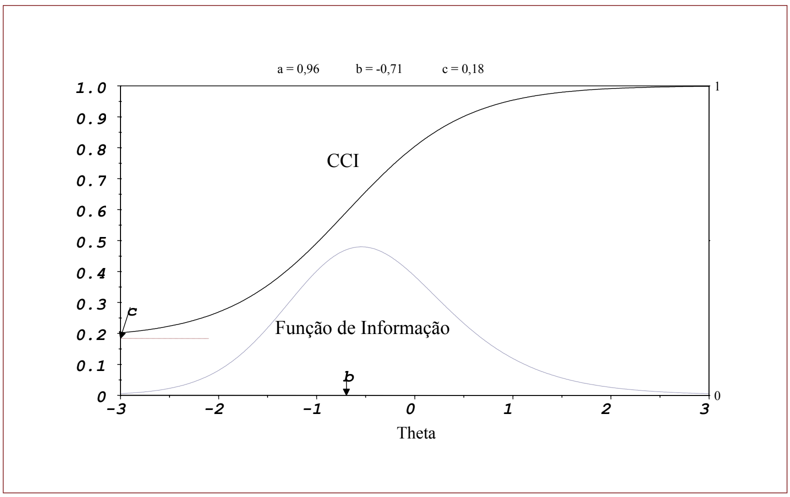
Com os parâmetros dos itens e o grau de aptidão dos indivíduos na variável latente medida pelo item (θ) tendo sido estabelecido, o investigador poderá, agora, determinar as funções de informação dos itens (FI) para inúmeros objetivos, dentre os quais estudar o funcionamento diferencial dos itens (DIF) ou organizar bancos de itens, conforme apregoa Andriola (1998, 2000, 2001, 2002) e Hambleton (1997).

Como o seu nome indica, a *função de informação (FI)* permite conhecer a contribuição do item à medida de θ, e o que é mais importante: em que ponto ou intervalo de θ a informação é máxima. Sua formulação é dada por $I_i(\theta) = \dfrac{[P_i^2(\theta)]}{P_i(\theta)Q_i(\theta)}$, na qual:

- $I_i(\theta)$ é a função de informação do item *i*, dado o valor θ;
- $P_i(\theta)$ é a probabilidade de acerto ao item *i*, dado θ;
- $Q_i(\theta)$ é o valor resultante de $1 - P_i(\theta)$;
- $P^2{}_i(\theta)$ é a derivada de $P_i(\theta)$.

Para o uso prático da FI, encontram-se representadas, na figura 2, a CCI e a respectiva FI de um hipotético item.

Figura 2 – Curva característica do item (CCI) e sua respectiva função de informação (FI)



Como se vê, a FI do item proporciona maior quantidade de informação no intervalo compreendido entre -1,5 ≤ θ ≤ 0,5 sendo, desse modo, mais útil à medida dessa variável latente para os sujeitos que tenham essa mesma magnitude em θ. Portanto, o item em foco permite que se cometa menos erros na estimação de θ no intervalo -1,5 ≤ θ ≤ 0,5.

Cabe ressaltar que, nos modelos logísticos de um e dois parâmetros, a informação dada pelos itens é máxima quando θ = b. No modelo de três parâmetros a informação é máxima quando $\theta = b + \left(\dfrac{1}{D_a}\right)\left\{\ln\left[\dfrac{1}{2} + \left(\dfrac{1}{2}\right)\sqrt{(1+8c)}\right]\right\}$, sendo o termo D_a uma constante de valor 1,7.

A quantidade de informação nesse ponto de θ, no qual o seu valor é máximo, é dada por $\frac{D^2}{4}$ para os modelos logísticos de um e dois parâmetros, e para o modelo logístico de três parâmetros é dada por $I(\theta)=\left[\frac{D^2a^2}{8(1-c)^2}\right]\left[1-20c-8c^2+(1+8c)^{\frac{3}{2}}\right]$.

Como ressaltamos, tem sentido conceitual a denominação *função de informação (FI),* pois quanto maior for I(θ) menor será o erro padrão de medida e, por conseguinte, maior a informação acerca de θ. Todavia, devemos enfatizar que as funções de informação dos itens (FI) e do teste (FT) dependem da escala na qual se encontrem expressados os valores de θ, conforme nos lembra Lord (1980). Dado que essa escala é arbitrária, o conceito e o valor da FI não são absolutos, depende, isso sim, da escala escolhida para medir θ. De acordo com o autor, isso não supõe um inconveniente grave, mas sua ignorância, por parte do avaliador ou do investigador, pode implicar conclusões equivocadas acerca da informação proporcionada pelo item ou pelo teste, nos diferentes níveis de θ.

CONSIDERAÇÕES FINAIS

A prática da avaliação psicológica e educacional sofreu profundas modificações com o surgimento dos modelos de TRI. Como vimos, desde então, foram propostos distintos modelos unidimensionais para dados dicotômicos; desenvolveram-se variados procedimentos para a estimação dos parâmetros dos itens e da magnitude dos sujeitos na variável latente θ – estimação de máxima verossimilhança condicional e estimação de máxima verossimilhança conjunta; formularam-se distintas provas para verificar o grau de ajuste do modelo aos dados empíricos – qui-quadrado e análise dos resíduos.

Nesse novo contexto, o foco das sofisticadas análises estatísticas desviou-se do teste (que representa o macro) para o item (que representa o micro). A lógica para tal câmbio é a mesma utilizada pela Física Moderna: *o micro exerce influência sobre o macro*. Justifica-se, desse modo, a acentuada ênfase no uso dos modelos de TRI, visto que se prestam à análise acurada do micro.

Nos Estados Unidos, os modelos unidimensionais de TRI são utilizados pelo reconhecido *National Assessment Educational Progress* (NAEP), pelas universidades em seus processos seletivos, pelo Exército e, ainda, pelo *Programme for International Student Assessment* (PISA). No Brasil, usados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), órgão subordinado ao Ministério da Educação e do Desporto (MEC), sobretudo no Sistema Nacional de Avaliação da

Educação Básica (Saeb). Ademais, algumas secretarias estaduais de educação, como a dos Estados de Minas Gerais e do Ceará, dentre outras, também utilizam a TRI como suporte estatístico para a análise de dados do sistema educacional (Barreto; Andriola, 2000; Soares; Genovez; Galvão, 2005).

Não obstante, o conhecimento acerca dos modelos unidimensionais de TRI, no âmbito brasileiro, é, ainda, bastante superficial. Isso é um problema que deverá ser encarado e combatido por meio da implementação de: cursos visando à formação de recursos humanos nas áreas da psicometria e da avaliação educacional; incentivo e incremento das investigações que usam a TRI; incentivo à publicação de *papers,* capítulos de livros e livros especializados, contendo relatos de experiências bem-sucedidas que tenham feito emprego da TRI; incentivo à divulgação de trabalhos acerca da TRI em congressos e eventos científicos de cunho nacional e/ou internacional.

Para concluir, desejamos ressaltar uma preocupação de caráter pessoal. Convém lembrar que os psicometristas, os psicólogos e os pedagogos manejam dados educacionais que dificilmente podem ser considerados unidimensionais. Assim, Hambleton (1997) e McDonald (1989) destacam que os modelos multidimensionais de TRI, empregados minoritariamente nos Estados Unidos e em alguns países da União Europeia, surgirão com muita força nos próximos anos, pois dados politômicos e multidimensionais têm maior ocorrência no âmbito educacional. Em outras palavras: *o modismo derivado do uso massivo dos modelos unidimensionais de TRI, iniciado nos anos 1980, está prestes a tornar-se procedimento ultrapassado*. Nesse contexto, nos ocorre citar o escritor, ator, cineasta e dramaturgo francês, Jean Cocteau (1889-1963) que, de modo bastante irônico, certa feita destacou: *a moda sempre morre jovem.*

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRIOLA, W. B. Descrição dos principais métodos para detectar o funcionamento diferencial dos itens DIF. *Psicologia:* reflexão e crítica, v. 14, n. 3, p. 643-652, 2001.

_________ . *Detección del funcionamiento diferencial del ítem DIF en tests de rendimiento:* aportaciones teóricas y metodológicas. 2002. 629 p. Tese (Doutorado) - Universidade Complutense de Madrid.

_________ . Principales métodos para la determinación del funcionamiento diferencial de los ítems DIF. In: CONGRESO NACIONAL, 12º, IBEROAMERICANO DE PEDAGOGÍA, 1º, 2000, Madrid. *Anais*... 2000. t. II, Resúmenes de Comunicaciones. p. 49-50.

_________ . Utilização da teoria da resposta ao item TRI para a organização de um banco de itens destinado à avaliação do raciocínio verbal. *Psicologia:* reflexão e crítica, v. 11, n. 3, p. 295-308, 1998.

ANDRIOLA, W. B.; BARRETO, J. A. E. Análise métrica de um instrumento de medida da aprendizagem através da teoria de resposta aos itens TRI. *Ensaio:* avaliação de políticas públicas em educação, v. 14, n. 5, p. 59-74, 1997.

ANSLEY, T. N.; FORSYTH, R. A. An Examination of the characteristics of unidimensional IRT parameters estimates derived from two dimensional data. *Applied Psychological Measurement*, v. 9, n. 1, p. 37-48, 1985.

BARRETO, J. A. E.; ANDRIOLA, W. B. O Mestrado em avaliação educacional da Universidade Federal do Ceará. In: BARRETO, J. A. E.; MOREIRA, R. V. O. (Org.). *Razão e fé do carvoeiro:* escritos de filosofia da ciência. Fortaleza: Programa Editorial Casa José de Alencar, 2000.

BERNSTEIN, I. H.; TENG, G. Factoring items and factoring scales are different: spurius evidence for multidimensionality due to item categorization. *Psychological Bulletin,* v. 105, p. 467-477, 1989.

CHRISTOFFERSSON, A. Factor analysis of dichotomized variables. *Psychometrika,* v. 40, n. 1, p. 5-32, 1975.

COSTA, L. F.; BIANCHI, A. G. C. A Outra dimensão da dimensão fractal. *Ciência Hoje,* v. 31, n. 183, p. 40-47, 2002.

CUESTA, M. Unidimensionalidad. In: MUÑIZ, J. (Org.). *Psicometría.* Madrid: Universitas, 1996.

DRASGOW, F.; PARSONS, C. K. Applications of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, n. 7, p. 189-199, 1983.

DUNHAM, W. *Viaje a través de los genios*: biografías y teoremas de los grandes matemáticos. Madrid: Pirâmide, 2002.

FERNANDES, T. Educação com qualidade. *Ciência Hoje,* v. 30, n.176, p. 56-58, 2001.

GARCÍA JIMÉNEZ, E.; GIL FLORES, J.; GÓMEZ, G. R. *Análisis factorial*. Madrid: Editorial la Muralla, 2000. (Cuadernos de estadística, 7).

GAVÍRIA SOTO, J. L. Fundamentos de la medida en educación y psicología: del representacionalismo a la medida como hipótesis. *Revista de Ciencias de la Educación*, n. 184, p. 207-223, out-dez., 2000.

__________ . *El Supuesto de la unidimensionalidad en la teoria del rasgo latente*: aportaciones metodológicas. Madrid: Editora da Universidad Complutense de Madrid, 1988.

GOLDSTEIN, H. Dimensionality, bias, independence and measurement scale problems in latent trait test score models. *British Journal of Mathematical and Statistical Psychology*, v. 33, p. 234-246, 1980.

GOLDSTEIN, H.; WOOD, R. Five decades of item response modelling. *British Journal of Mathematical and Statistical Psychology*, v. 42, p. 139-167, 1989.

HAIR JR., J. F. et al. *Multivariate data analysis*. Londres: Prentice Hall, 1999.

HAMBLETON, R. K. Aplication of item response models to criterion-referenced assessment. *Applied Psychological Measurement*, v. 7, n. 1, p. 33-44, 1983.

__________ . Item response theory: introduction and bibliography. *Psicothema,* v. 2, n. 1, p. 97-107, 1990.

__________ . Perspectivas futuras y aplicaciones. In: MUÑIZ, J. *Introducción a la teoría de respuestas a los ítems*. Madrid: Ediciones Psicologia Pirâmide, 1997. p. 203-213.

HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. *Fundamentals of item response theory.* Noth Caroline: Sage Publications, 1991.

HARRISON, D. A. Robustness of IRT parameter estimation to violations of the unidimensionality assumption. *Journal of Educational Statistics,* v. 11, n. 2, p. 91-115, 1986.

HATTIE, J. An Empirical study of various indices for determining unidimensionality. *Multivariate Behavioral Research*, n. 19, p. 49-78, 1984.

__________ . Methodology review: assessing unidimensionality of test and items. *Applied Psychologica Measurement,* v. 9, n.2, p. 139-164, 1985.

HATTIE, J. et al. An Assessment of stout's index of essential unidimensionality. *Applied Psychological Measurement,* v. 20, p. 1-14, 1996.

KEELING, K. B. A Regresion equation for determing the dimensionality of data. *Multivariate Behavioral Research,* v. 35, n.4, p. 457-46, 2000.

KNOL, D. L.; BERGER, M. P. F. Empirical comparasion between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, n. 26, p. 457-477, 1991.

LORD, F. M. *Applications of item response* *theory to pratical testing problems.* New Jersey: Lawrence Erlbaum Associates, 1980.

_________ . Maximum likelihood and bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, v. 23, n. 2, p. 157-162, 1986.

LORD, F. M.; NOVICK, M. *Statistical theories of mental tests scores.* Massachusets: Addison–Wesley, 1968.

MARTÍNEZ ARIAS, R. *Psicometria*: teoria de los tests psicológicos y educativos. Madrid: Ediciones Sintesis, 1997.

MCDONALD, R. P. Difficulty factors in binary data. *The British Journal of Mathematical and Statistical Psychology*, n. 27, p. 82-99, 1971.

_________ . The Dimensionality of test and items. *The British Journal of Mathematical and Statistical Psychology,* n. 33, p. 161-183, 1981.

_________ . Future directions for item response theory. *International Journal of Educational Research,* n. 13, p. 205-230, 1989.

__________. *Test Theory:* a unified treatment. New Jersey: Lawrence Erlbaum Associates, 1999.

MCDONALD, R. P.; AHLAWAT, K. S. Difficult factors in binary data. *Bristish Journal of Mathematical and Statistical Psychology*, n. 27, p. 82-99, 1974.

MISLEVY, R. J. Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistical,* n. 11, p. 3-31, 1986.

MUÑIZ, J. *Introducción a la teoria de respuesta a los ítems.* Madrid: Ediciones Pirâmide, 1997.

_________ . *Teoria clásica de los tests.* Madrid: Ediciones Pirâmide, 1994.

MUÑIZ, J.; HAMBLETON, R. K. Medio siglo de Teoría de Respuestas a los Ítems. *Anuario de Psicologia,* n. 52, p. 41-66, 1992.

MUTHÉN, B. Contributions to factor analysis of dichotomous variables. *Psychometrika*, v. 43, n. 4, p. 551-560, 1978.

_________ . Dichotomous factor analysis of sympton data. *Sociological Methods & Research,* n. 18, p. 19-65, 1989.

NANDAKUMAR, R. Assessing dimensionality of a set of items: comparisons of different approaches. *Journal of Educational Measurement*, n. 31, p. 17-35, 1994.

PASQUALI, L. *Psicometria*: teoria e aplicações. Brasília: UnB, 1997.

RECKASE, M. D. Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics,* v. 4, n. 3, p. 207-230, 1979.

ROUSSOS, L. A.; SCHNIPKE, D. L.; PASHLEY, P. J. A Generalized formula for the Mantel-Haenszel item functioning parameter. *Journal of Educational and Behavior Statistics,* v. 24, n. 3, p. 293-322, 1999.

SERRES, M. *Historia de las ciencias.* Madrid: Cátedra, 1998.

SOARES, T. M.; GENOVEZ, S. F. M.; GALVÃO, A. F. Análise do comportamento diferencial dos itens de geografia: estudo da 4ª série avaliada no Proeb/Simave 2001. *Estudos em Avaliação Educacional*, v. 16, n. 32, p. 81-110, 2005.

STOUT, W. et al. DIMTEST: a fortran program for assessing dimensionality of binary items responses. *Applied Psychological Measurement,* v. 16, p. 236, 1992.

VAN DER LINDEN, W. J.; HAMBLETON, R. K. *Handbook of modern item response theory.* New York: Springer Verlag, 1997.

WALLER, N. G. *MicroFACT 1.0. A microcomputer factor analysis program for dichotomous and ordered polytomous data and mainframe sized problems.* Illinois: Assessment Systems Corporation, 1995.

WHERRY, R.; GAYLORD, R. H. Factor pattern of test items and tests as a function of the correlation coefficient: content, difficult and constant error factor. *Psychometrika,* v. 9, p. 237-244, 1944.

Recebido em: fevereiro 2009
Aprovado para publicação em: julho 2009