

Comportamento Diferencial do Item - (DIF): uma apresentação

RAQUEL DA CUNHA VALLE

Doutoranda em Estatística - Universidade de São Paulo/USP

Estatístico do Departamento de Pesquisas Educacionais da

Fundação Carlos Chagas, São Paulo - SP

rvalle@fcc.org.br

Resumo

O artigo procura suprir uma deficiência na literatura nacional que trata de modelos estatísticos aplicados à pesquisa e, especialmente, à avaliação educacional. O uso intensivo da Teoria da Resposta ao Item apresenta dificuldades sobretudo para pessoas sem uma formação matemática mais aprofundada. Assim, nas avaliações desenvolvidas a partir dos anos 90 aparecem certas siglas e termos, como, por exemplo, *DIF - Differential Item Functioning*, que oferecem dificuldades para alguns educadores. O presente trabalho procura explicar, na medida do possível, sem aprofundamento matemático, o conceito de *Comportamento Diferencial do Item*, largamente empregado em avaliação educacional, inclusive no Brasil.

Palavra-chave: Teoria da Resposta ao Item, Comportamento Diferencial do Item, Análise de Itens.

Resumen

El artículo tiene por objetivo suprir una deficiencia en la literatura nacional referente a modelos estadísticos aplicados en la pesquisa y, en especial, en la evaluación educacional. El uso intensivo de la Teoría de la Respuesta al Ítem tiene presentado dificultades en especial para las personas sin un conocimiento matemático más profundo. De ese modo, surgen ciertas siglas y términos en las evaluaciones desarrolladas a partir de los años 90, como por ejemplo, *DIF - Differential Item Functioning*, que ofrecen dificultades para algunos educadores. Este trabajo intenta explicar, en lo posible, sin profundidad matemática, el concepto de *Comportamiento Diferencial del Ítem*, largamente aplicado en la evaluación educacional, incluso en Brasil.

Palabras-clave: Teoría de la Respuesta al Ítem, Comportamiento Diferencial del Ítem, Análise de Itens.

Abstract

This article attempts to fill a gap in the Brazilian literature on statistical models applied to research and, especially, to educational evaluation. The intensive use of Item Response Theory is especially difficult for people without adequate mathematical training. In the evaluations developed since the nineties certain acronyms and terms come up, e.g. *DIF - Differential Item Functioning*, which are difficult for certain educators. This article attempts to explain, as far as possible, and without going in any depth into mathematical terms, the concept of *Differential Item Functioning*, widely used in education evaluation, including in Brazil.

Keywords: Item Response Theory, Differential Item Functioning, Item Analysis.

Introdução

Com o aumento da utilização de técnicas derivadas da Teoria da Resposta ao Item-TRI (ver, por exemplo, Hambleton, Swaminathan e Rogers, 1991), na área de Avaliações Educacionais, cresce também o número de pessoas ligadas a essa área que desconhece boa parte dos termos e das técnicas estatísticas mais utilizadas. Talvez esse problema se deva, em grande parte, à falta de textos básicos, direcionados aos educadores e a outros profissionais da área que não tenham, necessariamente, formação matemática.

Em muitos dos textos mais recentes sobre avaliações aparecem referências ao DIF - *Differential Item Functioning* (em português, comportamento diferencial do item), mas apenas pequena parte dos leitores deve realmente saber do que se trata. Este trabalho procura explicar, de maneira bastante introdutória e sem muito aprofundamento matemático, o que é o DIF. Para tanto, utilizaremos alguns resultados encontrados na análise DIF feita pelo SAEB - Sistema de Avaliação da Educação Básica.

Começando com uma definição, uma das mais utilizadas é: *um item mostra DIF se indivíduos com a mesma habilidade, mas de diferentes grupos, não têm a mesma probabilidade de acerto no item* (Hambleton et. al., 1991). Inicialmente, vamos procurar entender os conceitos envolvidos nessa definição. Usaremos o termo habilidade para nos referirmos aos traços latentes do indivíduo numa dada área de conhecimento, ou seja, características do indivíduo que não podem ser observadas diretamente.

Na Teoria Clássica das Medidas (ver, por exemplo, Vianna, 1987) usualmente trabalhamos com o escore dos indivíduos, que é uma medida facilmente observada a partir da aplicação de um teste e, nesse contexto, consideramos sempre a prova como um todo. Na TRI, utilizamos as respostas do indivíduo a cada item e com base nessa informação, utilizamos modelos matemáticos convenientes que nos permitem estimar a habilidade do indivíduo na área de conhecimento desejada. Muitos textos também utilizam o termo proficiência como sinônimo da habilidade.

Uma das grandes vantagens da TRI sobre a Teoria Clássica é que ela nos permite fazer comparações que antes só seriam possíveis se todos os indivíduos fossem submetidos às mesmas provas ou pelo menos a formas paralelas de testes. Por exemplo, a TRI viabiliza a comparação entre populações, desde que submetidas a provas que tenham alguns itens comuns ou, ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente distintas. Isto porque

uma das principais características da TRI é que ela tem como elementos centrais os itens, e não a prova como um todo.

Dessa maneira, várias questões de interesse prático na área da Educação podem ser respondidas. É possível, por exemplo, avaliar o desenvolvimento de uma determinada série de um ano para outro ou comparar o desempenho entre escolas públicas e privadas. Mas, obviamente, a comparação de uma teoria com a outra não resulta somente em vantagens. As ferramentas matemáticas utilizadas na TRI são bem mais complexas, exigindo uma análise estatística mais sofisticada, além de programas computacionais específicos. Como consequência disso, talvez a maior desvantagem da utilização da TRI seja o distanciamento que se pode criar entre quem processa os cálculos e quem efetivamente vai utilizar esses resultados, dificultando a compreensão e a interpretação adequada das informações fornecidas.

Voltando aos conceitos utilizados na definição do DIF, é também importante definir os conceitos de grupo e população, da maneira como serão considerados nesse trabalho. É comum que uma população seja definida por determinadas características que podem variar, dependendo dos objetivos do estudo e, portanto, podem ou não ser relevantes para a diferenciação de uma população de outra, dependendo do caso. Por exemplo, pode-se considerar que a 4.^a série do Ensino Fundamental de São Paulo é a população alvo. Daí, toma-se uma amostra dos alunos dessa população. Já em outro estudo, poderíamos considerar a 4.^a série do Ensino Fundamental do Brasil como a população de interesse. Então, seria tomada uma amostra contemplando os vários estados do Brasil. Assim, São Paulo juntamente com todos os demais estados seria considerada uma população. Exemplos do que usualmente são consideradas como populações distintas são séries distintas (3.^a série e 4.^a série); períodos distintos (diurno e noturno); uma mesma série, mas em anos distintos (4.^a série de 1999 e 4.^a série de 2001), etc.

Já quando usarmos o termo grupo, nesse texto, estaremos nos referindo a subdivisões de uma mesma população. Por exemplo, considerando a variável sexo, teremos 2 grupos (meninos e meninas), dentro da população da 4.^a série do Ensino Fundamental. Outras variáveis freqüentemente utilizadas para definir grupos de interesse podem ser raça, região de procedência, etc.

No Brasil, uma das mais importantes avaliações do ensino básico, que possui alcance nacional e tem servido de modelo para as demais avaliações é o SAEB. Nessa avaliação, a análise do DIF é feita por regiões, ou seja, os grupos considerados são as cinco grandes regiões do Brasil: Norte, Nordeste, Sul, Sudeste e Centro-Oeste.

Para que serve a análise do DIF?

O objetivo de uma análise de DIF é investigar se um item tem ou não o mesmo comportamento para indivíduos pertencentes a dois grupos distintos, mas de mesma habilidade, ou seja, verificar basicamente se um item apresenta graus de dificuldade diferentes para subgrupos da população que têm o mesmo nível de conhecimento.

Parece claro que, em princípio, não seja desejável encontrar itens que apresentem DIF, pois isso significaria, num primeiro momento, que algum particular grupo de indivíduos estaria sendo privilegiado em detrimento dos demais. No entanto, nem sempre itens diagnosticados com DIF são considerados "ruins" e nem sempre o procedimento adotado é simplesmente retirá-los das provas. Isto porque a análise do DIF pode ser uma ferramenta de diagnóstico do sistema educacional bastante útil.

Por exemplo, pode-se utilizar essa análise para a detecção de conteúdos que são abordados com ênfases diferentes ou para a detecção de diferenças regionais de currículo. O resultado da análise DIF aponta em quais regiões o ensino de alguns temas precisa ser enfatizado, não havendo nesses casos razão para eliminar tais itens das provas. Outra estratégia que também pode ser adotada é reexaminar os itens com problemas de DIF em função dos objetivos da prova. Assim, tais itens podem ser reformulados e reaproveitados.

Falando especificamente de Brasil, "no SAEB a vantagem imediata da análise DIF é proporcionar a identificação de itens que tratam de temas com conotações regionais ou possíveis diferenças de ênfases curriculares, mas que abordam assuntos reconhecidamente de interesse nacional. Por exemplo, em Geografia, os tipos de vegetação de cada região; em História, fatos ocorridos em um determinado Estado, como a Inconfidência Mineira, a Guerra de Canudos ou a Revolta de Farroupilha; em Português, o uso de expressões locais e gírias, etc"¹.

O que pode causar o DIF?

Grande parte das causas determinantes do DIF ainda são desconhecidas. Estudos conduzidos pelo *Educational Testing Service* - ETS, nos Estados Unidos, apontam que o DIF pode ser causado basicamente por uma trinca de fatores: a **familiaridade** com o conteúdo do item, também

¹ Sistema Nacional de Avaliação da Educação Básica: SAEB 99 - Relatório da análise de comportamento diferencial dos itens (Differential Item Functioning - DIF) entre regiões (2000).. A partir desse Relatório serão apresentados alguns exemplos típicos para fins de esclarecimento do conceito de DIF.

referida como exposição ao tema ou fator cultural; o interesse pessoal naquele dado conteúdo e a reação emocional negativa provocada pelo conteúdo.

No caso específico do Brasil, "tendo em vista a dimensão continental do país e as peculiaridades de cada uma de suas regiões que, certamente se refletem na vida social, econômica e cultural da população, incluindo nela a educação e suas formas de dinamização nas escolas, o comportamento de um item pode diferir porque, em seu enunciado, ilustrações e alternativas de respostas aparecem":

- temas regionais, mais familiares em determinadas região do que em outras;
- características lingüísticas, como termos, expressões e gírias locais usados em algumas regiões, mas não em todas;
- fatos ocorridos em um estado/região e, portanto, nele mais conhecidos;
- nomes/palavras que associam a resposta certa do item a algum aspecto específico da região;
- temas provavelmente mais focalizados pelo ensino de uma região;
- temas que possivelmente não são igualmente explorados no currículo das cinco regiões por diferença de ênfase"².

Métodos matemáticos para detecção de DIF

Para entendermos melhor as idéias dos métodos utilizados na detecção do DIF, vamos novamente voltar a alguns dos conceitos básicos da Teoria da Resposta ao Item. A TRI propõe a utilização de modelos que representam a probabilidade de um indivíduo responder corretamente a um item como função dos parâmetros do item e da habilidade do respondente. Essa relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item.

Dentre os modelos propostos pela TRI, o modelo logístico de três parâmetros é atualmente um dos mais utilizados. É dado por:

$$P(X_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-D a_i(\theta_j - b_i)}}$$

com $i = 1, 2, \dots, I$ itens e $j = 1, 2, \dots, n$ indivíduos, onde:

² Sistema Nacional de Avaliação da Educação Básica: SAEB 99 - Relatório da análise de comportamento diferencial dos itens (Differential Item Functioning - DIF) entre regiões (2000)..

X_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente ao item i , ou 0 quando o indivíduo j não responde corretamente ao item i .

θ_j representa a habilidade (traço latente) do j -ésimo indivíduo.

$P(X_{ij} = 1 | \theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i .

b_i é o parâmetro de dificuldade (ou de posição) do item i , medido na mesma escala da habilidade.

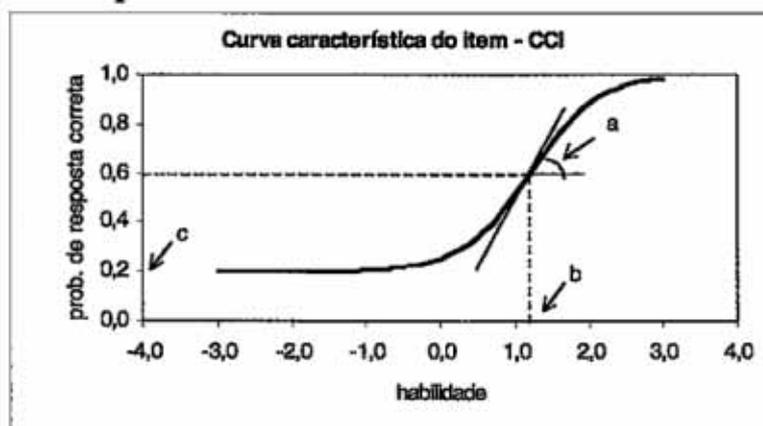
a_i é o parâmetro de discriminação (ou de inclinação) do item i .

c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente ao item i (muitas vezes referido como a probabilidade de acerto casual).

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

Note que $P(X_{ij} = 1 | \theta_j)$ pode ser vista como a proporção de respostas corretas ao item i dentre todos os indivíduos da população com habilidade θ_j . A relação existente entre $P(X_{ij} = 1 | \theta_j)$ e os parâmetros do modelo é mostrada na figura a seguir, que é chamada de Curva Característica do Item (CCI).

Figura 1
Exemplo de uma Curva Característica do Item - CCI



Como pela definição um item apresenta DIF se sua CCI não é a mesma para grupos diferentes. Podemos, então, pensar em investigar o DIF comparando as CCI de dois ou mais grupos e isto pode ser feito de várias maneiras. A primeira, e talvez mais intuitiva, seria comparar os parâmetros que determinam a CCI (que no caso do modelo logístico de três parâmetros seriam o a , b e c definidos acima).

Existem várias críticas a esse método. Uma delas é que o teste estatístico utilizado na comparação dos parâmetros é assintótico, ou seja, a distribuição da estatística do teste não é exata, o que significa, em linhas gerais, que sua distribuição exata só seria conhecida no caso de grandes amostras. No entanto, pode não ser apropriado o uso de distribuições assintóticas no caso em que os parâmetros de item (os parâmetros a , b e c) e as habilidades (os parâmetros θ) são estimados simultaneamente, ou seja, são todos desconhecidos, como é a situação mais freqüente.

Outra crítica ao método é que diferenças significativas entre os parâmetros podem ser encontradas quando, na prática, não existem diferenças na faixa de variação da habilidade de mais interesse. A escala de habilidade é arbitrária, mas é comum considerarmos por exemplo, uma escala com média 0 e desvio padrão 1. Nesse tipo de escala (que foi considerada na ilustração da CCI da Figura 1), a amplitude de variação de interesse concentra-se de -3 a 3. Considerando os parâmetros das CCI's de 2 grupos como:

Grupo 1:	$a = 1.8$	$b = 3.5$	$c = 0.2$
Grupo 2:	$a = 0.5$	$b = 5.0$	$c = 0.2$

estudos mostraram que foram encontradas diferenças estatísticas significativas entre os parâmetros das CCI's dos dois grupos e, no entanto, as duas curvas diferem menos do que 0.05 na região de variação de habilidade de maior interesse (-3 a 3).

Uma outra maneira seria comparar as CCI's avaliando a área abaixo da curva definida por cada uma delas. Por este método, calcula-se a diferença entre as áreas sob as duas curvas. Foram desenvolvidos expressões exatas para o cálculo dessas diferenças quando o parâmetro c pode ser considerado como sendo o mesmo para os dois grupos.

Quando o parâmetro c não é o mesmo para ambos os grupos e considerando-se uma variação finita na escala de habilidades (por exemplo, de -3 a 3), métodos numéricos devem ser usados para o cálculo dessas diferenças. No entanto, nesse caso existem problemas para determinar o valor crítico que irá definir qual a magnitude da diferença entre as duas curvas que deve ser considerada estatisticamente significativa.

Devido aos problemas associados a esses e também a outros métodos de detecção do DIF associados a TRI, métodos alternativos foram propostos. Tais métodos não utilizam técnicas derivadas da TRI na detecção do DIF, ou seja, não se baseiam nos parâmetros de itens nem em suas CCI's. Dentre os mais conhecidos podemos citar o método de Mantel-Haenszel (Holland e Thayer, 1988) e o procedimento de detecção do DIF através de regressão logística, proposto por Swaminathan e Rogers (1990).

Considerando-se os itens corrigidos de maneira dicotômica (certo ou errado), existem 2 tipos diferentes de DIF: (a) o DIF uniforme, que ocorre quando o item favorece uniformemente um grupo em relação a outro e (b) o DIF não-uniforme, quando há uma interação entre o nível de habilidade e a performance no item, de modo que a direção do DIF muda ao longo da escala de habilidade. Uma crítica ao método de Mantel-Haenszel (MH) é que ele não é sensível ao DIF não-uniforme. Este problema motivou a busca por técnicas de detecção do DIF que superassem essa limitação, como é o caso da regressão logística. No entanto, o método MH ainda é a metodologia mais utilizada para análise do DIF, inclusive pelo *Educational Testing Service*, nos exames do *National Assessment for Educational Progress* (NAEP), e aqui no Brasil, na análise do SAEB. Sendo assim, vamos explicar mais detalhadamente este procedimento.

Uma terminologia bastante empregada nos textos que tratam da análise do DIF é tratar os dois grupos em estudo como grupo de referência, que é o grupo tomado como padrão, e grupo focal, que é o grupo no qual encontra-se o interesse da análise DIF.

No método MH, uma tabela $2 \times 2 \times K$ com os dados dos respondentes é construída com base na performance (acerto ou erro) dos membros de cada grupo (focal e de referência) e o escore total no teste. Ou seja, este procedimento emparelha os grupos em uma medida de desempenho, que nas aplicações usuais é o escore total no teste. Assim, para cada um dos K níveis da variável de emparelhamento, o procedimento MH constrói uma tabela 2×2 , como a tabela abaixo:

Tabela 1
Divisão dos respondentes no procedimento de Mantel-Haenszel

Grupo	Desempenho no Item		total
	acerto	erro	
referência	a_k	b_k	n_{Rk}
focal	c_k	d_k	n_{Fk}
total	m_{1k}	m_{0k}	T_k

Assim, para um dado escore k , n_{Rk} e n_{Fk} são, respectivamente, o número de alunos nos grupos de referência e focal, m_{1k} é o número de alunos que acertou o item, m_{0k} é o número de alunos que errou o item e T_k é o número total de alunos.

A razão de chances MH estimada é então usada para comparar os dois grupos em termos da sua chance de responder ao item corretamente, condicionada ao escore total no teste. O procedimento MH utiliza a hipótese de que essa razão de chances α é constante em todos os K níveis da variável de emparelhamento. Assim, a estatística $\hat{\alpha}_{MH}$ é dada por:

$$\hat{\alpha}_{MH} = \frac{\sum_{k=1}^K a_k d_k / T_k}{\sum_{k=1}^K b_k c_k / T_k}$$

Em geral, utiliza-se uma transformação dessa a estatística, dada por:

$$\hat{\Delta}_{MH} = -2.35 \log_e(\hat{\alpha}_{MH})$$

Por convenção, essa transformação é definida de modo que esta medida assume valores negativos quando o item é mais difícil (condicionalmente aos valores da variável de emparelhamento) para o grupo focal.

O erro padrão estimado para esta estatística é:

$$EP(\hat{\Delta}_{MH}) = 2.35 \sqrt{\text{Var}(\log_e(\hat{\alpha}_{MH}))} \quad \text{onde}$$

$$\text{Var}(\log_e(\hat{\alpha}_{MH})) = \frac{\sum_{k=1}^K u_k v_k / T_k^2}{2 \left(\sum_{k=1}^K a_k d_k / T_k \right)^2} \quad \text{com}$$

$$u_k = (a_k d_k) + \hat{\alpha}_{MH} (b_k c_k)$$

e

$$v_k = (a_k + d_k) + \hat{\alpha}_{MH} (b_k + c_k)$$

No caso específico do SAEB, que utiliza um planejamento de itens do tipo Blocos Balanceados Incompletos (BBI), onde um mesmo item aparece em vários cadernos, é necessária a utilização de um procedimento mais complexo, proposto por Allen e Donoghue (1996). Esta abordagem tem a vantagem de produzir somente uma análise MH para cada item, levando em conta a informação contida em cada caderno em que o item aparece.

Um caso particular: O DRIFT

O DRIFT (tendência) é um estudo de DIF quando os grupos de indivíduos considerados na análise são indexados pela variável tempo. O objetivo é verificar se um mesmo item tem comportamento diferente para alunos de mesma habilidade porém em tempos diferentes. O DRIFT pode ocorrer em testes educacionais nos quais há itens comuns ao longo dos anos. Isto pode ocorrer devido a possíveis mudanças no currículo ou à exposição de determinados temas na mídia.

Exemplos de itens com DIF encontrados no SAEB

Na análise do SAEB 99 foram identificados vários itens com DIF nas diferentes séries e disciplinas avaliadas pelo estudo. O estudo do DIF foi feito por regiões, isto é, os grupos foram definidos pelas 5 grandes regiões brasileiras: Norte, Nordeste, Sudeste, Sul e Centro-Oeste. A análise foi feita de forma a se considerar inicialmente a região Norte como grupo de referência em relação às outras. A SEGUIR, UTILIZOU-SE A REGIÃO Nordeste como grupo de referência em relação às demais, e assim sucessivamente.

Para uma melhor interpretação dos valores de $\hat{\Delta}_{MH}$ que serão apresentados nos exemplos, devemos lembrar que $\hat{\Delta}_{MH}$ é uma estatística cujo numerador está relacionado aos acertos do grupo de referência e erros do grupo focal, e cujo denominador ao complementar dessas quantidades, ou seja, aos erros do grupo de referência e acertos do grupo focal. Dessa maneira, essa estatística fornece uma medida do desempenho do grupo de referência com relação ao grupo focal. Valores de $\hat{\Delta}_{MH}$ iguais a 1 significam que os dois grupos tiveram o mesmo desempenho no item; valores de $\hat{\Delta}_{MH} > 1$, indicam um melhor desempenho do grupo de referência em relação ao grupo focal e, finalmente, valores de $\hat{\Delta}_{MH}$ entre 0 e 1, indicam um melhor desempenho do grupo focal. Portanto, pela definição matemática de $\hat{\Delta}_{MH}$

apresentada anteriormente, teremos $\hat{\Delta}_{MH}$ negativo quando o desempenho do grupo de referência no item é melhor do que o do grupo focal e $\hat{\Delta}_{MH}$ positivo caso contrário, como se encontra resumido no quadro abaixo:

$\hat{\sigma}_{MH}$	$\hat{\Delta}_{MH}$	Melhor desempenho no item
$0 < \hat{\sigma}_{MH} < 1$	$\hat{\Delta}_{MH} > 0$	Grupo focal
$\hat{\sigma}_{MH} = 1$	$\hat{\Delta}_{MH} = 0$	Mesmo desempenho nos 2 grupos
$\hat{\sigma}_{MH} > 1$	$\hat{\Delta}_{MH} < 0$	Grupo de referência

Além do sinal de $\hat{\Delta}_{MH}$ também existe uma relação direta entre sua magnitude e a diferença entre o desempenho dos dois grupos de maneira que, quanto maiores essas diferenças, maior o valor de $\hat{\Delta}_{MH}$. O quadro a seguir fornece alguns exemplos da relação entre $\hat{\sigma}_{MH}$ e $\hat{\Delta}_{MH}$:

$\hat{\sigma}_{MH}$	$\hat{\Delta}_{MH}$
1	0
2	1.629
5	3.782
10	5.411
100	10.822
1000	16.233

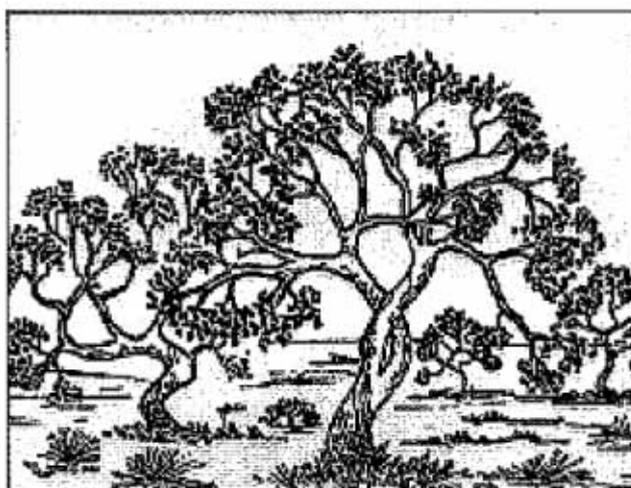
Cabe apenas ressaltar que o quadro anterior exemplifica situações onde o grupo de referência tem igual ou melhor desempenho que o grupo focal ($\hat{\sigma}_{MH} \geq 1$); no entanto, no caso de situações inversas $\hat{\Delta}_{MH}$ terá o mesmo valor com sinal negativo. Exemplificando, no caso de $\hat{\sigma}_{MH} = 1/5 = 0.2$, teremos $\hat{\Delta}_{MH} = -3.782$.

A seguir apresentamos um exemplo de item com DIF em cada uma das disciplinas avaliadas na 4ª série do Ensino Fundamental do SAEB.

Geografia

O item a seguir, que ilustrava a vegetação do cerrado, típica da região Centro-Oeste, apresentou DIF positivo entre esta região e todas as outras, sendo portanto, mais fácil para os alunos do Centro-Oeste.

Observe a gravura:



A formação vegetal retratada na gravura é:

- (A) campo.
- (B) cerrado.
- (C) floresta.
- (D) vegetação de montanha.

Fonte: Ministério da Educação e do Desporto (2000). Sistema Nacional de Avaliação da Educação Básica: SAEB 99 - Relatório da análise de comportamento diferencial dos itens (Differential Item Functioning - DIF) entre regiões. Rio de Janeiro: Fundação Cesgranrio, p. 13.

Resultados da análise DIF:

Grupo de REFERÊNCIA	Grupo FOCAL	$\hat{\Delta}_{MH}$
Sul	Centro-Oeste	3.310
Sudeste	Centro-Oeste	3.104
Nordeste	Centro-Oeste	2.760
Norte	Centro-Oeste	2.536

História

Este é outro exemplo de um item mais fácil para alunos da região Centro-Oeste do que para alunos das regiões Norte e Sudeste, de mesma proficiência. Talvez isso se deva ao fato de que o texto faça referência aos Estados de Mato Grosso e Goiás, onde moram os alunos da região Centro-Oeste.

Leia atentamente este texto:

No início do século XVIII, os bandeirantes partiram para o interior, descobriram minas de ouro nos atuais Estados de Mato Grosso e Goiás, alargando as fronteiras do território brasileiro para além dos limites estabelecidos pelo Tratado de Tordesilhas.



Que região do Brasil foi ocupada, no século XVIII, com a prática da mineração dos bandeirantes paulistas?

- (A) Centro - Oeste
- (B) Nordeste
- (C) Norte
- (D) Sudeste

Fonte: Ministério da Educação e do Desporto (2000). Sistema Nacional de Avaliação da Educação Básica: SAEB 99 - Relatório da análise de comportamento diferencial dos itens (Differential Item Functioning - DIF) entre regiões. Rio de Janeiro: Fundação Cesgranrio, p. 43.

Resultados da análise DIF:

Grupo de REFERÊNCIA	Grupo FOCAL	$\hat{\Delta}_{MH}$
Sudeste	Centro-Oeste	2.138
Norte	Centro-Oeste	1.723

Língua Portuguesa

Este item, que aborda a Festa de Parintins, típica da região Norte, obteve DIF favorável a alunos dessa região em relação aos alunos das regiões Sudeste, Sul e Centro-Oeste, de mesma habilidade. Este item também mostrou-se mais fácil para alunos da região Nordeste do que para alunos da região Sul.

Carnaval à moda da selva

Parintins, no coração da Amazônia, explode em cores e paixões para comemorar o boi-bumbá, uma festança de rara beleza que encanta os turistas.

É noite na floresta. A 420 quilômetros de Manaus, pelas águas do Rio Amazonas, um caldeirão azul e vermelho, com o formato estilizado de um boi, acende em Parintins. Tudo ao redor está escuro. Tum-tum-tum. Os surdos começam a marcar o ritmo. Entra na arena a marujada de guerra, a bateria do boi Caprichoso. As 15 mil pessoas que estão de azul cantam, gritam, agitam-se como as águas de uma pororoca. Na metade de lá das arquibancadas, a galera de vermelho, que torce para o outro boi, o Garantido, mantém-se num silêncio amazônico. É o início da festa do boi-bumbá, realizada todos os anos, entre 28 e 30 de junho, no coração da selva. Nas três noites de espetáculo, os dois bois revezam-se no bumbódromo. E, num respeito assustador, há sempre silêncio de uma turma quando o adversário está evoluindo. Terminada a festa, o suspense: quem vencerá a disputa? O resultado foi anunciado: deu Caprichoso na cabeça. Apesar dos descontentes, a folia pede passagem e garante mais um dia de animação.

É quase um milagre que, numa cidade pobre como Parintins, com renda per capita de apenas um salário mínimo, se realize o magistral espetáculo de cor, luz e som do boi-bumbá.

ÉPOCA, Rio de Janeiro: Globo, n. 7, 6 julho, 1998. (Fragmento)

Releia as expressões abaixo:

"o boi Caprichoso"

"o boi Garantido"

"uma pororoca"

"o bumbódromo"

Este vocabulário é específico da Região:

- (A) Sul.
- (B) Norte.
- (C) Leste.
- (D) Centro-Oeste.

Fonte: Ministério da Educação e do Desporto (2000). Sistema Nacional de Avaliação da Educação Básica: SAEB 99 - Relatório da análise de comportamento diferencial dos itens (Differential Item Functioning - DIF) entre regiões. Rio de Janeiro: Fundação Cesgranrio, p. 62.

Resultados da análise DIF:

Grupo de REFERÊNCIA	Grupo FOCAL	$\hat{\Delta}_{MH}$
Norte	Centro-Oeste	-1.931
Norte	Sudeste	-2.018
Nordeste	Sul	-2.019
Norte	Sul	-2.790

Matemática

Diferentemente das disciplinas já analisadas, os DIF's em Matemática não podem ser explicados pelo caráter de regionalismo como em História e em Geografia, nem pelas variações lingüísticas de regiões ou classes sociais, como em Português. Acredita-se que os DIF encontrados em Matemática se devam à maior ou menor ênfase dada a certos procedimentos de ensino em determinadas regiões e à priorização de certos temas no currículo.

O item abaixo, por exemplo, mostrou-se mais fácil para os alunos das regiões Sudeste e Sul do que para alunos da região Nordeste, de mesma proficiência.

Ao realizar um cálculo, Lucas obteve o seguinte resultado: 287.
Que cálculo Lucas pode ter realizado?

(A) $2 + 8 + 7$
 (B) $200 + 80 + 7$
 (C) $208 + 7$
 (D) $20 + 87$

Fonte: Ministério da Educação e do Desporto (2000). Sistema Nacional de Avaliação da Educação Básica: SAEB 99 - Relatório da análise de comportamento diferencial dos itens (Differential Item Functioning - DIF) entre regiões. Rio de Janeiro: Fundação Cesgranrio, p. 69.

Resultados da análise DIF:

Grupo de REFERÊNCIA	Grupo FOCAL	$\hat{\Delta}_{MH}$
Nordeste	Sudeste	1.979
Nordeste	Sul	1.764

Ciências

Este item apresentou um DIF favorável para alunos da região Norte em relação aos do Sudeste, Nordeste e Sul, de mesma proficiência. Pode-se supor que este resultado seja explicado por um maior conhecimento do jabuti na região Norte.

Alguns animais podem mudar de cor para se confundir com o ambiente e outros utilizam parte do próprio corpo para se defenderem dos predadores. Qual é a defesa do jabuti?

- (A) Os dentes.
- (B) As garras.
- (C) A carapaça.
- (D) O ferrão.

Fonte: Ministério da Educação e do Desporto (2000). *Sistema Nacional de Avaliação da Educação Básica: SAEB 99 - Relatório da análise de comportamento diferencial dos itens (Differential Item Functioning - DIF) entre regiões*. Rio de Janeiro: Fundação Cesgranrio, p. 83.

Resultados da análise DIF:

Grupo de REFERÊNCIA	Grupo FOCAL	$\hat{\Delta}_{MH}$
Norte	Sudeste	-1.663
Norte	Nordeste	-1.806
Norte	Sul	-2.147

Referências bibliográficas

ALLEN, N. L. & DONOGHUE, J. R. (1996). Applying the Mantel-Haenszel Procedure to Complex Samples of Items. *Journal of Educational Measurement*, 33, p. 231-251.

BAKER, F. B. (1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.

FRENCH, A. W. & MILLER, T. R. (1996). Logistic Regression and Its Use In Detecting Differential Item Functioning in Polytomous Items. *Journal of Educational Measurement*, 33, p. 315-332.

GULLIKSEN, H. (1950). *Theory of Mental Tests*. New York: John Wiley and Sons.

- HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. (1991). **Fundamentals of Item Response Theory**. Newbury Park: Sage Publications.
- HOLLAND, P. W.; THAYER, D. T. (1988). Differential item Performance and the Mantel-Haenszel procedure. In H. Wainer & H. Braun (Eds.), **Test Validity** (p. 129-145) Hillsdale, NJ: Erlbaum
- LINDEN, W. J. van der; HAMBLETON, R. K. (1997). **Handbook of Modern Item Response Theory**. New York: Springer-Verlag.
- LORD, F. M.; NOVICK, M. R. (1968). **Statistical Theories of Mental Test Score**. Reading: Addison-Wesley.
- MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO (2000). **Sistema Nacional de Avaliação da Educação Básica: SAEB 99 - Relatório da análise de comportamento diferencial dos itens (Differential Item Functioning - DIF) entre regiões**. Rio de Janeiro: Fundação Cesgranrio.
- STRICKER, L. J.; EMMERICH, W. (1999). Possible Determinants of Differential Item Functioning: Familiarity, Interest and Emotional Reaction. **Journal of Educational Measurement**, 36, p. 347-366.
- VALLE, R. C. (1999). **Teoria da Resposta ao Item**. Dissertação de Mestrado apresentada ao Departamento de Matemática e Estatística da Universidade de São Paulo.
- VIANNA, H. M. (1987). **Testes em Educação**. São Paulo: IBRASA
- ZWICK, R.; THAYER, D. T.; LEWIS, C. (1999). An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. **Journal of Educational Measurement**, 36, p. 1-28.

