

O Funcionamento Diferencial do Item (DIF) como estratégia para captar ênfases curriculares diferenciadas em matemática

GLAUCO DA SILVA AGUIAR*

RESUMO

O texto apresenta o *Differential Item Functioning* (DIF) como uma importante ferramenta estatística na identificação de diferenças em testes de habilidade cognitiva que fazem uso da Teoria de Resposta ao Item (TRI), a exemplo do que ocorre nas avaliações educacionais em larga escala. Um item apresenta DIF quando alunos de diferentes grupos que possuem a mesma habilidade cognitiva não têm a mesma probabilidade de acertarem o item. O presente trabalho utiliza os resultados do Brasil e de Portugal, no Programa Internacional de Avaliação dos Estudantes (PISA), de 2003, cuja área de conhecimento privilegiada foi a Matemática. Foram analisados 84 itens aplicados a alunos de 41 países, sendo 4452 brasileiros e 4608, portugueses. Os resultados discutidos mostram que alguns itens apresentam funcionamento diferencial entre alunos brasileiros e portugueses, principalmente em função da subárea da Matemática e do contexto em que o conhecimento matemático é aplicado.

Palavras-chave: Funcionamento Diferencial do Item (DIF), Teoria de Resposta ao Item (TRI), Ensino de Matemática, PISA.

RESUMEN

El texto presenta el *Differential Item Functioning* (DIF) como una importante herramienta estadística para la identificación de diferencias en tests de habilidades cognitivas que usan

* Professor do Colégio Militar do Rio de Janeiro (glaucoaguiar@uol.com.br).

la Teoría de la Respuesta al Ítem (TRI), como ejemplo de lo que ocurre en las evaluaciones educativas a gran escala. Un Ítem presenta DIF cuando alumnos de diferentes grupos que poseen la misma habilidad cognitiva no tienen la misma probabilidad de acertar el ítem. El presente trabajo utiliza los resultados de Brasil y de Portugal, en el Programa Internacional de Evaluación de los Estudiantes (PISA), del 2003, cuya área de conocimiento privilegiada fue la Matemática. Se analizaron 84 ítems aplicados a alumnos de 41 países, de los cuales 4452 eran brasileños y 4608, portugueses. Los resultados discutidos muestran que algunos ítems presentan un funcionamiento diferencial entre alumnos brasileños y portugueses, principalmente en función de la subárea de matemática y del contexto en el que se aplica el conocimiento matemático.

Palabras clave: Funcionamiento Diferencial del Ítem (DIF), Teoría de Respuesta al Ítem (TRI), Enseñanza de Matemática, PISA.

ABSTRACT

This article presents DIF – Differential Item Functioning – as an important statistical tool in identifying differences in tests of cognitive ability that make use of IRT– Item Response Theory –, similar to what occurs in educational assessments on a large scale. An item is said to present a differential functioning when students from different groups, who have the same cognitive ability, do not have the same probability of answering the item correctly. This study was developed using Brazil's and Portugal's results in the Program for International Student Assessment (PISA) in 2003, whose privileged area of assessment was Mathematics. We analyzed 84 Mathematics items administered to students from 41 countries, out of which 4452 were Brazilian students and 4608, Portuguese ones. The results discussed here show that some Mathematics items present differential functioning between Brazilian and Portuguese students mainly because of the subarea of Mathematics and the context to which mathematical knowledge is applied.

Keywords: Differential Item Functioning (DIF), Item Response Theory (IRT), Mathematics Education, PISA.

INTRODUÇÃO

Ao avaliar as competências e habilidades, ao fim da escolarização básica, o *Programme for International Student Assessment* (PISA), uma pesquisa internacional realizada pela Organização para Cooperação e Desenvolvimento Econômico (OCDE) examina o grau de preparação dos jovens de 15 anos de idade para a vida adulta e, até certo ponto, a efetividade dos sistemas educacionais de diferentes países. Essa efetividade refere-se às realizações dos objetivos subjacentes dos sistemas educacionais, como definidos pela sociedade, ou seja, investigar até que ponto está sendo levada a efeito “uma visão do Ensino Médio de caráter amplo, de forma que os aspectos e conteúdos tecnológicos associados ao aprendizado científico e matemático sejam parte essencial da formação cidadã de um sentido universal, e não somente de sentido profissionalizante” (Brasil, 1999). Essa matriz conceitual de avaliação do PISA, que tem como referência principal a articulação entre o conceito de educação básica e o de cidadania, encontra-se em sintonia com os objetivos educacionais do Ensino Médio, já propostos pelo Ministério da Educação (MEC), e está presente também no Exame Nacional do Ensino Médio (Enem) cujos preceitos demonstram seu caráter transdisciplinar e o entendimento do perfil terminativo desse nível de ensino, complementando o aprendizado iniciado no ensino fundamental.

No entanto, desde o primeiro ciclo de avaliação, realizado em 2000, em função do desempenho insatisfatório dos alunos brasileiros, a divulgação dos resultados tem como foco as conclusões enfáticas de que, em termos educacionais, o Brasil não está bem. Os professores ensinam mal e os alunos aprendem cada vez menos. Estudos comparativos de sistemas educacionais, porém, não devem se limitar apenas a medir e comparar os resultados educacionais brutos, conseguidos pelos alunos, mas recorrer a metodologias que possibilitem conjugar e articular os paradigmas quantitativo e qualitativo, a fim de identificar os principais fatores capazes de explicar as diferenças de rendimento encontradas e analisar o modo como interagem entre si (Ferrer, 2003).

Nessa perspectiva, meu objetivo foi identificar fatores capazes de explicar as diferenças de rendimentos encontradas no letramento em Matemática, entre alunos brasileiros e seus colegas portugueses. Considerando que os sistemas diferem entre si, com uma variação que se pode considerar natural, e que as características próprias que os distinguem têm consequências nos diversos modos de elaboração e desenvolvimento do currículo; e, ainda, que os conteúdos são selecionados pelos professores e abordados com ênfases diferenciadas, procurei identificar as características dos itens de teste, em relação à forma de apresentação, aos contextos em que

são aplicados os conteúdos e às áreas da Matemática que sinalizassem a existência de ênfases curriculares diferenciadas, nesses dois países. Sendo a análise de DIF (*Differential Item Functioning*) uma ferramenta estatística que possibilita extrair dos resultados dos testes esses padrões de efeitos diferenciados, a análise consistiu na aplicação de métodos para detectar e identificar os itens que apresentaram funcionamento diferencial entre o Brasil e Portugal.

De acordo com Soares, Genovez e Galvão (2005), para possibilitar a comparabilidade dos resultados, é essencial que o modelo utilizado na avaliação educacional garanta o pressuposto de que o item apresente o mesmo “funcionamento” para os diversos grupos populacionais que estão sendo avaliados. No caso de modelos da TRI, isto significa manter a estabilidade dos parâmetros dos modelos dos itens para as diferentes populações. Para uma boa comparação entre resultados de grupos diferentes de alunos, é imprescindível, pois, uma atenção especial à construção dos itens, a fim de que estes não apresentem funcionamento diferencial. Um item apresenta DIF entre dois ou mais grupos distintos, quando estes forem agrupados, sistematicamente, em grupos de mesma habilidade cognitiva, e, mesmo assim, as probabilidades de acerto do item forem significativamente diferentes para os grupos pareados. Assim, na estimação das proficiências, o ideal é evitar o emprego de itens com DIF elevado, isto é, os que favoreçam um determinado grupo de alunos, em detrimento de outros.

Embora o DIF possa significar que algum grupo particular de indivíduos esteja sendo privilegiado, em detrimento dos demais, ainda assim, a evidência estatística da diferença entre o desempenho de grupos não deve ser persuasiva quanto à decisão de excluir ou não um item do teste, pois sua análise pode ser uma ferramenta de diagnóstico do sistema educacional bastante útil. Doolittle e Cleary (1987) mostraram que o desempenho das meninas é inferior ao dos meninos, quando se trata de itens que medem habilidade matemática em geometria e em raciocínio matemático. Tais habilidades são objetivos legítimos na Educação Matemática e os estudantes, sejam meninos, sejam meninas, devem saber lidar igualmente com problemas dessa natureza. Excluir tais itens de um teste, por favorecerem a determinado grupo, torna o instrumento incompleto e é prejudicial para os que estão em desvantagem, pois tende a perpetuar a diferença.

O FUNCIONAMENTO DIFERENCIAL DO ITEM (DIF)

Estudos visando a identificar itens que sejam favoráveis a determinado grupo, em detrimento de outros, ganham destaque no campo da psicometria moderna,

pois ajudam a assegurar que os testes sejam tão imparciais quanto é possível fazê-los. Nesse sentido, Soares, Genovez e Galvão (2005) destacam que a preocupação com o funcionamento diferencial do item antecede ou, ainda, extrapola o contexto da TRI, em que a ausência do DIF é requisito para uma boa equalização entre resultados de grupos diferentes de alunos.

Historicamente, a preocupação com o DIF está fortemente associada ao desejo de que se construíssem questões de teste que não fossem afetadas por características étnico-culturais dos grupos submetidos aos testes de avaliação educacional (Cole, 1993) – muito ligada, portanto, às campanhas em prol da melhoria dos direitos civis dos cidadãos comuns, nos anos de 1960, nos Estados Unidos da América. Esses anos foram marcados por uma enorme preocupação com a igualdade de oportunidades, pelas críticas aos sistemas educacionais discriminadores, pelo desenvolvimento de um conceito popular e legal de ações afirmativas e pela consciência racial/étnica. Diferenças educacionais, resultantes de sistemas educacionais com muita iniquidade, passaram a ser vistas como vestígios de uma velha ordem segregadora. Assim, escores de testes, refletindo essas diferenças, foram considerados, da mesma forma, discriminadores, e passou-se a usar o termo *viés*, ao referenciá-los.

Iniciam-se, assim, estimulados pela discussão social, alheia, em grande parte, ao círculo psicométrico, estudos para desenvolver formas de identificação do *viés*, nos itens e nos testes. Esses estudos tinham por objetivo provar que os testes ou instrumentos de medida não possuíam nenhum tipo de *viés* (Cole, 1993). Então, sob a seguinte concepção de *viés*: um item é enviesado se sujeitos de habilidades iguais, mas de culturas diferentes, não têm a mesma probabilidade de acertar o item (Angoff, 1973; Linn et al., 1981; Shepard, Camilli, Averill, 1993; Ironson, 1993; Linn, Drasgow, 1993), muitos pesquisadores começaram a se dedicar ao estudo sistemático das diferenças entre os grupos étnicos, com o objetivo de tentarem encontrar explicações convincentes para as grandes diferenças de rendimento, observadas entre os diversos grupos étnicos e socioeconômicos, que refletiam, na realidade, disparidades nas oportunidades educacionais e se mostravam injustos, ao exigirem tarefas estranhas às culturas de algumas minorias.

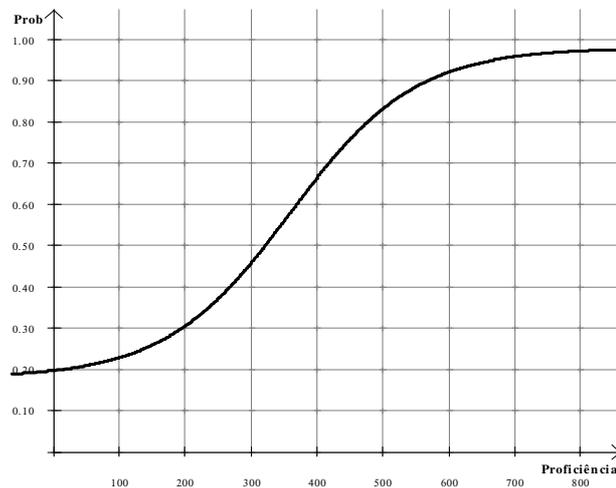
No âmbito da TRI, é possível dizer que o item não tem DIF, quando a curva característica do item (CCI) é a mesma para os grupos comparados em um mesmo nível de habilidade ou proficiência (θ) medida pelo item. Em linguagem matemática, podemos expressar a ausência de DIF com respeito à variável G (grupo) dado Z (nível de θ) se, e somente se, $F(X | g, z) = F(X | z)$, onde:

- X é a pontuação no item ($X=1$ correto, $X=0$ errado);
- g é o valor obtido, segundo a variável G ;
- z é o valor obtido, segundo a variável Z .

Nesse contexto, os valores esperados por $E(X | g, \theta) = E(X | \theta)$, para todo o g e θ . No caso de itens dicotômicos, os valores esperados são as probabilidades de acerto ao item, que podem ser expressas nos seguintes termos:

$P(X = 1 | g, \theta) = P(X = 1 | \theta)$, para todo g e θ . No segundo caso, $P(X = 1 | \theta)$, a equação expressa, na realidade, a curva característica do item (CCI), representada na figura abaixo (Andriola, 2006).

Figura 1 – Curva característica do item (CCI)



Dentre os modelos propostos pela TRI, um dos mais utilizados é o modelo logístico de três parâmetros, cuja equação é dada por:

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

Com $i = 1, 2, 3, \dots, I$ itens e $j = 1, 2, 3, \dots, n$ indivíduos, onde:

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 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 ao item i .

b_i é o parâmetro de dificuldade do item, medido na mesma escala da habilidade.

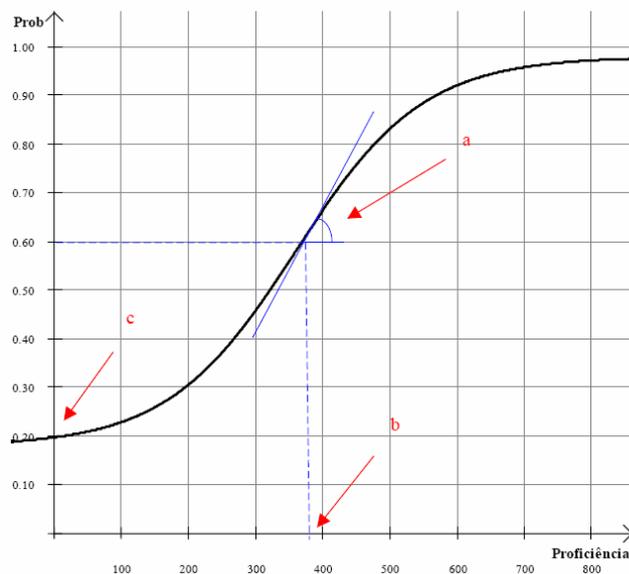
a_i é o parâmetro de discriminação (ou de inclinação) do item i . Refere-se à capacidade de o item distinguir alunos com diferentes níveis de habilidade.

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

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1.7, quando se deseja que a função logística forneça resultados semelhantes aos 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 apresentada na figura abaixo, chamada de Curva Característica do Item (CCI).

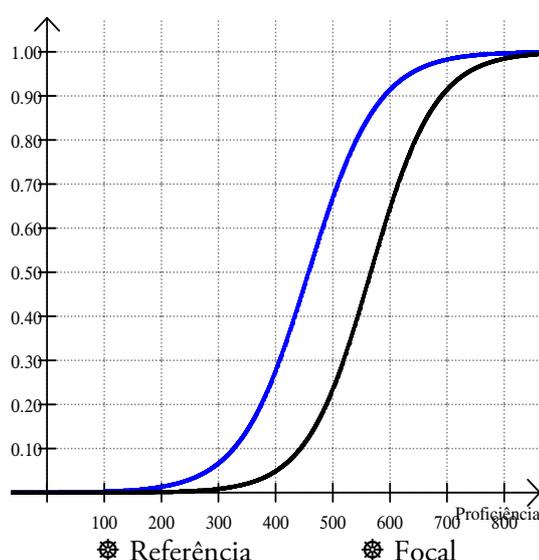
Figura 2 – Modelo logístico de três parâmetros



Em conformidade com o pressuposto da TRI de que a probabilidade de acerto ao item é função da proficiência do aluno, essa curva tem que ser a mesma, para dois grupos de alunos que tenham a mesma proficiência. Dito de outra maneira, um item apresenta DIF, portanto, se sua CCI não é a mesma para grupos diferentes, no nosso caso, para países diferentes.

Existem, basicamente, dois tipos de DIF. O primeiro é o DIF *uniforme* ou *consistente*, que ocorre quando as CCIs do item estudado para o Grupo de Referência e para o Grupo Focal são diferentes, indicando que o item favorece uniformemente um dos grupos, em relação ao outro. Em outras palavras, as curvas não se cruzam em nenhum ponto, ao longo da proficiência (θ). A figura abaixo mostra um exemplo de item, apresentando DIF uniforme.

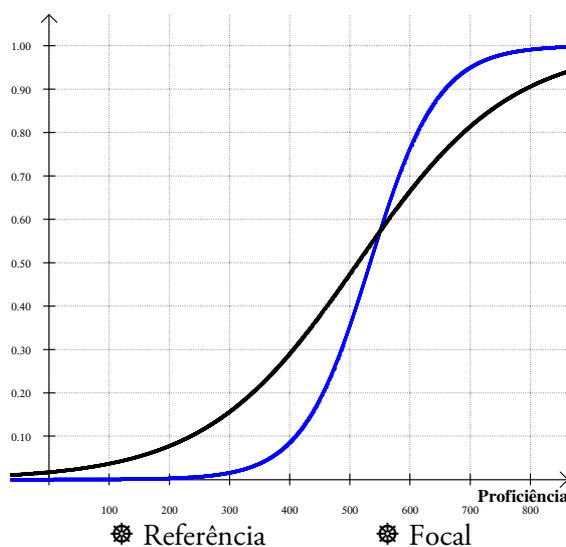
Figura 3 – Representação de um item com DIF uniforme



De acordo com a figura 3, observa-se que a CCI do grupo de referência está situada mais à esquerda que a CCI do grupo focal, o que indica que o item é mais fácil para o grupo de referência, em todos os níveis de proficiência. Essa diferença aponta que o item apresenta DIF, nesse caso, favorável ao grupo de referência. Supondo que as curvas representam dois países, cujos respectivos alunos foram submetidos a um mesmo item, poderíamos afirmar que esse item apresenta DIF no parâmetro **b**, ou seja, apenas na dificuldade. Isto porque o parâmetro **c** é igual a zero, para os dois grupos e a inclinação da curva, descrita pelo parâmetro **a**, é também a mesma, para os dois grupos. De acordo com esse exemplo, alunos com proficiências iguais a 500, nos dois grupos, têm chances diferentes de acertarem o item. O grupo focal tem 25% e o grupo de referência, 65%, o que caracteriza um comportamento anômalo desse item.

O segundo tipo de DIF, denominado DIF *não-uniforme* ou *inconsistente*, ocorre quando há uma interação entre o nível de proficiência e a performance no item, de modo que a direção do DIF muda, ao longo da escala de proficiência. Observa-se que as CCIs são diferentes e se cruzam, em algum ponto do contínuo da proficiência, como pode ser observado na figura 4.

Figura 4 – Representação de um item com DIF não-uniforme



Fonte: Relatório Técnico do PISA 2003.

Assim, de acordo com esse exemplo, para níveis de proficiências mais baixas, o item favorece o grupo focal. À medida que temos os dois grupos nivelados por proficiências mais altas, o DIF se inverte e passa a favorecer o grupo de referência.

MÉTODOS DE INVESTIGAÇÃO DE DIF

Existem vários procedimentos formais, para se estudar itens com DIF. De modo geral, aqueles podem ser divididos em dois grupos: os chamados clássicos, que necessitam de uma proficiência já conhecida, e os métodos baseados nos modelos da TRI, que não precisam de uma proficiência já conhecida, mas que dependem de alguma hipótese que garanta a comparabilidade dos resultados de proficiência, para os grupos analisados. E, em particular, que exista e seja conhecido, *a priori*, um subconjunto de itens que não possuam DIF.

Uma revisão dos métodos clássicos pode ser encontrada, por exemplo, em Andriola (2002), Soares, Genovez e Galvão (2005) e Valle (2002).

Um significativo desenvolvimento nessa área foi alcançado com o artigo de Lord e Novick (1968), no qual explicam o modelo da Teoria de Resposta ao Item (TRI). Logo, ficou evidente que esse modelo poderia ser usado, com proveito, no estudo do funcionamento diferencial do item. Como se sabe, agora, a base da teoria reside na função da resposta ao item, ou seja, a curva em forma de S, da proporção de indivíduos de mesmo nível de habilidade, que responde corretamente a um determinado item. Pressupondo que a habilidade considerada seja unidimensional e que o item meça essa habilidade, a curva é única, sob as condições de um modelo particular; exceto para variações aleatórias, a mesma curva é encontrada, independentemente da natureza do grupo para o qual a função é plotada. A curva é frequentemente definida por três parâmetros: **a**, **b** e **c**, como apresentados na figura 2. Em razão da natureza única da curva de resposta ao item, sob as condições mencionadas, o fato de a curva de resposta não ser a mesma para dois grupos é a evidência de que os pressupostos não são satisfatórios para um ou ambos os grupos. Podemos, então, pensar em investigar a presença de DIF, comparando os parâmetros que determinam a CCI.

A fim de superar os problemas associados a esses métodos e também a outros derivados da TRI, foram desenvolvidos métodos alternativos que não utilizam técnicas derivadas da TRI, na detecção do DIF, ou seja, não-paramétricos. Dentre os mais conhecidos, estão o procedimento de detecção de DIF por meio da regressão logística, proposto por Swaminathan e Rogers (1990) e o Método de Mantel-Haenszel (Holland, Thayer, 1993). Este é o mais utilizado para a 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 (Valle, 2002).

METODOLOGIA APLICADA

Ao dar início a este estudo, objetivando identificar os itens que apresentaram DIF, vali-me do método da regressão logística, adotando como conhecida a proficiência estimada no PISA. A variável **pv1math** (*Plausible value in math*), presente na base de dados do PISA e calculada para os examinandos dos diferentes países, foi utilizada no modelo de regressão para a estimação do parâmetro **b** de dificuldade dos itens. A partir dos b_i encontrados, foi possível investigar o DIF por intermédio da diferença entre esses parâmetros, para os dois países considerados. Depois de calculados os valores acima e ter-se verificado que alguns itens apresentavam DIF considerável,

foram produzidas as CCI, a fim de ratificar o comportamento anômalo de alguns itens. Esses gráficos foram produzidos, utilizando o *software* matemático winplot¹ e os valores de β_0 e β_1 estimados na regressão logística. Posteriormente, para obter a magnitude do DIF encontrado, usei os *softwares* BILOG-MG 3.0² e SisAni (Sistema de Análise de Itens)³ para o cálculo da estatística de Mantel-Haenszel (MaH), cujos valores servem para identificar se um item apresenta DIF e qual a sua magnitude. Os resultados dessas análises indicaram que, dentre os 84 itens da prova de Matemática do PISA 2003, 23 apresentam algum tipo de funcionamento diferencial, sendo 11 deles a favor do Brasil e 12, a favor de Portugal. Esses itens que apresentaram DIF e respectivas direções (pró Brasil ou pró Portugal) e magnitudes constam do apêndice.

Na sequência, recorrendo a um item público, liberado para divulgação pelo consórcio que administra o PISA⁴, exemplifico, resumidamente, esses procedimentos adotados na identificação do DIF. Esclareço que o mesmo se deu para os demais 84 itens do PISA 2003, mas estes não serão apresentados aqui.

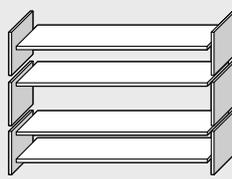
- O item a seguir (M484Q01), foi um dos 84 analisados:

ESTANTES

Questão 1: ESTANTES M484Q01

Para construir uma estante completa, um marceneiro precisa do seguinte material:

4 pranchas grandes de madeira,
6 pranchas pequenas de madeira,
12 braçadeiras pequenas,
2 braçadeiras grandes e
14 parafusos.



O marceneiro possui em estoque 26 pranchas grandes de madeiras, 33 pranchas pequenas de madeira, 200 braçadeiras pequenas, 20 braçadeiras grandes e 510 parafusos.

Quantas estantes completas o marceneiro poderá fazer?

Resposta:

¹ Disponível em: <http://math.exeter.edu/rparris/winplot.html>.

² Software estatístico utilizado na estimação de modelos da TRI (Zimowski et al., 2006).

³ Software estatístico desenvolvido por pesquisadores do Centro de Políticas Públicas e Avaliação da Educação (CAEd) da Universidade Federal de Juiz de Fora (UFJF).

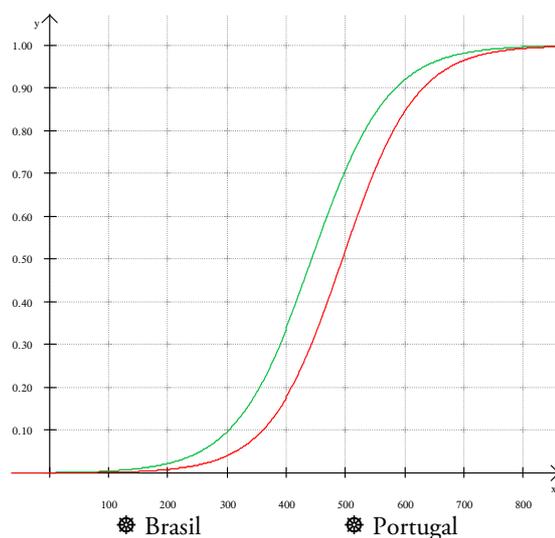
⁴ Os demais itens públicos utilizados na avaliação estão disponibilizados em: www.inep.gov.br.

Após calculados os parâmetros de dificuldade desse item para os dois grupos, a diferença entre os valores foi de 50,033 e o erro padrão, de 0,088, mostrando que esse item foi mais fácil para o Brasil ($b_P - b_B > 0$).

Com os valores de β_0 e β_1 estimados na regressão logística, produzi as CCI's do item, para os dois grupos. Essa análise gráfica, que normalmente é feita como primeiro passo para a verificação da qualidade dos itens, possibilitando verificar e avaliar, por exemplo, o grau de dificuldade e o poder de discriminação do item, mostrou-se bastante eficiente, também, na detecção do DIF. Como a CCI não é a mesma para os dois grupos, isso indica que o item possui algum tipo de funcionamento diferencial.

Abaixo, é apresentada a CCI desse item, para o Brasil e Portugal.

Gráfico 1 – Curva característica do item M484Q01T



O gráfico 1 nos mostra que esse item apresenta DIF apenas na dificuldade (parâmetro **b**), pois as curvas têm a mesma inclinação, indicando a não existência de DIF na discriminação (parâmetro **a**). Para todas as faixas de proficiência, o item é mais fácil para o Brasil. O valor da estatística delta de Mantel Haenszel, calculado para esse item (alfa D MH = -1,914), utilizada para analisar a magnitude do DIF encontrado, classifica-o como um DIF de magnitude alta. O sinal negativo indica que o item favorece o grupo de referência (Brasil).

Os gráficos a seguir, produzidos pelos *softwares* BILOG-MG e SisAni, dão uma ideia do ajuste do modelo aos dados empíricos e de como se comporta o item, em

Este resultado é coerente com os anteriores: diferença nos parâmetros de dificuldade e análise feita por via da curva característica do item – CCI.

Concluída essa fase do estudo, o passo seguinte foi buscar identificar algum padrão ou informação adicional que pudesse estar associada à existência do DIF e que fosse relevante para que se entendesse algumas das possíveis diferenças educacionais existentes entre os países analisados.

RESULTADOS E DISCUSSÕES DAS ANÁLISES DE DIF COM OS DADOS DO PISA

Uma vez identificados os itens que se mostraram mais fáceis para os estudantes brasileiros, ou para os alunos portugueses, o objetivo seguinte foi encontrar uma explicação para esse fato. Saliento que, no âmbito deste trabalho, a expressão “itens mais fáceis para alunos de um grupo em relação a outro grupo” pressupõe, sempre, comparações controladas pela proficiência em Matemática. Em outras palavras, comparo alunos com desempenhos semelhantes no teste.

A operacionalização para se alcançar esse objetivo foi realizada com base nas características dos itens, definidas pelo próprio PISA, como: “Subárea da Matemática”, “Processo”, “Contexto” e “Tipo de Resposta”, e duas outras características, definidas por mim, na expectativa de que pudessem ajudar na explicação do DIF. São elas: “Tamanho do Enunciado” e “Recurso Gráfico”, no caso de o item vir acompanhado de alguma figura, gráfico ou tabela.

A distribuição dos 23 itens de Matemática que apresentaram algum DIF, tendo o Brasil como grupo de referência e Portugal como grupo focal, consta na tabela 1. A classificação dos itens, para fins de explicação do DIF, é feita, neste momento, em função da subárea da Matemática que o item avalia.

Tabela 1 – Itens com DIF, segundo a subárea da matemática

Característica do Item (subárea)	Nº de itens com DIF Mais fácil para	
	Brasil	Portugal
Quantidade	7	3
Mudança e Relações	1	4
Espaço e Forma	1	2
Incerteza	2	3
Total	11	12

Observemos que as diferenças entre a quantidade de itens para cada grupo, dentro das subáreas, são maiores, em duas delas: “Quantidade” e “Mudança e Relações”. Dos dez itens da subárea “Quantidade”, sete foram mais fáceis para alunos brasileiros. Quanto à “Mudança e Relações”, dos cinco itens que apresentaram DIF, quatro foram mais fáceis para alunos portugueses. Esses dados indicam, num primeiro momento, que alunos brasileiros tendem a ter melhor desempenho do que alunos portugueses, quando o conteúdo avaliado refere-se à “Quantidade”, quando comparado com a subárea “Mudança e Relações”. Tal fato aponta um possível caminho, na tentativa de encontrar um padrão para explicar o DIF entre esses dois grupos. O êxito nesse empreendimento passa, necessariamente, por uma análise mais detalhada das características desses itens.

Classificando os mesmos 23 itens que apresentaram DIF, em razão das diferentes situações em que os estudantes encontram problemas matemáticos, ou, ainda, nas situações em que são aplicados os conhecimentos relevantes, encontrei quatro, os quais são de contexto “Científico” e tendem a apresentar DIF, beneficiando o grupo focal (Portugal), indicando, portanto, serem mais fáceis para alunos portugueses. Por outro lado, quando o contexto do item é “Pessoal”, os itens que apresentam DIF tendem a beneficiar o grupo de referência (Brasil).

Tabela 2 – Itens com DIF, segundo o contexto

Característica do Item (contexto)	Nº de itens com DIF Mais fácil para	
	Brasil	Portugal
Científico	0	4
Pessoal	4	0
Educacional/Ocupacional	3	5
Público/Social	4	3
Total	11	12

A classificação dos itens, em função dos processos da Matemática que têm de ser desempenhados pelos estudantes na resolução dos itens, é apresentada na tabela a seguir. Nota-se que as diferenças entre a quantidade de itens para cada grupo, nos diferentes processos, são maiores para “Reprodução” e “Conexão”.

Tabela 3 – Itens com DIF, segundo o processo

Característica do Item (processo)	Nº de itens com DIF Mais fácil para	
	Brasil	Portugal
Reprodução	2	6
Conexão	5	2
Reflexão	4	4
Total	11	12

A dimensão do processo define as capacidades necessárias para a Matemática que, no caso da “Reprodução”, restringe-se basicamente a operações matemáticas simples. Dos oito itens que apresentaram DIF, seis foram favoráveis aos alunos portugueses. Por outro lado, em relação à “Conexão”, que se afina com a capacidade de o aluno associar ideias para resolver problemas, dos sete itens que apresentaram DIF, cinco foram favoráveis ao Brasil.

Por fim, classificando os itens em função do tamanho do enunciado, encontrei o seguinte quadro:

Tabela 4 – Itens com DIF, segundo o tamanho do enunciado

Característica do Item (enunciado)	Nº de itens com DIF Mais fácil para	
	Brasil	Portugal
Curto	4	4
Médio	3	8
Longo	4	0
Total	12	12

Observa-se que, dentre os itens que apresentaram DIF, quatro deles têm a característica de possuírem um “Enunciado Longo”. Em todos esses itens, o resultado foi favorável ao Brasil, ou seja, mostraram-se mais fáceis para alunos brasileiros, em comparação aos seus colegas portugueses de mesma habilidade cognitiva. Num primeiro momento, esse resultado pode-nos causar certa estranheza, por contrariar resultados e diagnósticos anteriores, os quais revelam que nossos alunos leem mal os textos que lhes são apresentados e deles conseguem extrair

poucas informações para uma reflexão posterior. No entanto, ao analisarmos detalhadamente esses itens, percebemos que o enunciado longo, na verdade, descreve instruções a serem seguidas, as quais vêm acompanhadas de exemplos ou ilustrações, indicando a tarefa a ser cumprida. Apenas um desses itens não apresenta diretamente os passos a serem seguidos, acompanhado de exemplos, mas, implicitamente, nele está a ideia dos procedimentos a seguir. Provavelmente, para os alunos brasileiros, esse tipo de enunciado seja mais inteligível, e as descrições dos passos a serem seguidos ajudem na concentração do que está sendo pedido e facilite a resolução do problema.

A classificação dos 23 itens, segundo as características “Tipo de Resposta” e “Recurso Gráfico”, não apresentou diferenças significativas nessa fase exploratória, visando a um padrão para explicar o DIF. No entanto, analisando os itens mais cuidadosamente, verifiquei que aqueles que requeriam a interpretação de diferentes tipos de gráficos eram favoráveis aos alunos portugueses.

RESUMO DAS HIPÓTESES FORMULADAS PARA A EXPLICAÇÃO DO DIF

Explicar o DIF não é tarefa fácil. Além disso, toda estrutura técnica e pedagógica de uma avaliação da envergadura do PISA é empregada no sentido de se construírem itens de qualidade, que não apresentem DIF entre os diferentes grupos. No entanto, em função das características próprias de cada país, seu grau de desenvolvimento econômico e consequente influência na vida social e cultural de sua população, com reflexos na educação, alguns itens apresentaram DIF. O desafio na tentativa de explicá-los é encontrar, nos itens que favorecem determinado país, a existência de padrões. Para isso, o ideal seria que houvesse grande número de itens bastante diferenciados entre si, o que não foi o caso deste estudo. Um número maior de itens nesse tipo de análise talvez pudesse apontar para padrões mais claros e definidos.

No entanto, a análise descritiva das características dos itens de Matemática que apresentaram funcionamento diferencial, entre os alunos brasileiros e portugueses, mostrou-se bastante interessante, revelando a existência de diferenças relacionadas à *performance* do item entre esses grupos e apontando caminhos para identificar padrões que estariam causando o DIF entre os grupos.

No quadro 1, sintetizo, após análise detalhada dos itens, os possíveis padrões encontrados para explicar o DIF entre alunos brasileiros e portugueses. Ressalto, ainda, que, da forma como foi organizado o quadro, os itens que apresentam os padrões descritos tendem a apresentar DIF favorável ao país em questão.

Quadro 1 – Resumo das características dos itens com DIF

Grupos	Padrões
Brasil	<ul style="list-style-type: none"> • Subárea: Quantidade; • Contexto: Pessoal; • Processo: Conexão; • Itens da subárea Quantidade, envolvendo operações algébricas básicas com números inteiros; • Enunciado dos itens com orientações passo a passo e descrevendo procedimentos a serem seguidos para a resolução do item.
Portugal	<ul style="list-style-type: none"> • Subárea: Mudança e Relações; • Contexto: Científico; • Processo: Reprodução; • Itens da Subárea Quantidade, envolvendo operações algébricas básicas com representação decimal; • Itens envolvendo interpretação de variados tipos de gráficos.

Essas hipóteses formuladas para a explicação do DIF, descritas acima, carecem de uma confirmação mais rigorosa e formal, que pode ser obtida a partir de uma estrutura de regressão imposta à explicação da magnitude do DIF, e que tenha como co-variáveis independentes a indicação dos itens que trazem na sua forma e conteúdo as características referidas acima. Soares (2007) propõe um novo modelo da TRI: uma generalização do modelo logístico de três parâmetros, que incorpora a detecção de itens com DIF em sua estrutura. Nesse modelo, além dos parâmetros **a** (discriminação), **b** (dificuldade) e **c** (acerto casual), são incluídos dois parâmetros – d_{ig}^a e d_{ig}^b – para representarem o funcionamento diferencial na discriminação e na dificuldade, respectivamente, do item *i* no grupo *g*. O modelo proposto tem a estrutura a seguir:

$$P(Y_{ij} = 1 | \theta_j, a_i, b_i, c_i, d_{ig}^a, d_{ig}^b) = c_i + (1 - c_i) \frac{1}{1 + e^{-De^{d_{ig}^a} a_i (\theta_j - b_i + d_{ig}^b)}}$$

para $i = 1, 2, \dots, I$ $j = 1, 2, \dots, J$ e $g = 1, 2, \dots, G$

A utilização desse modelo integrado tem como objetivos confirmar ou não as hipóteses levantadas e verificar a concordância entre os itens detectados com DIF, por meio das abordagens tradicionais, e os detectados com DIF, pela abordagem integrada. Esse modelo foi utilizado neste estudo e os resultados corroboram as hipóteses formuladas que emergiram da análise exploratória⁵.

⁵ Maiores detalhes sobre esses resultados ver Aguiar (2008).

CONCLUSÃO

As características próprias de cada país exercem influência na organização do seu sistema educacional, na forma como os currículos são organizados e na ênfase com que se explora determinado conteúdo em sala de aula, entre outras. Essas características interferem no desempenho dos alunos, por razões que extrapolam, per si, a proficiência de cada um. A possibilidade de se conhecer essas características por meio dos itens que favorecem determinados grupos e de perceber a existência de padrões que passam despercebidos aos olhos dos especialistas que os elaboram é, sem dúvida, a grande contribuição que a análise do DIF traz para a avaliação educacional. Estudos dessa natureza evidenciam que uma boa comparabilidade das proficiências de diferentes grupos de alunos depende de que haja itens comuns, aplicados a esses grupos, que não apresentem DIF. Contudo, a quase impossível tarefa de se construir instrumentos de avaliação sem DIF, para grupos muito diferentes, pode ser revertida a favor do avaliador. O DIF pode ajudar a conhecer as diferenças, e esse é o primeiro passo para as ações educacionais que visam a eliminá-las ou, pelo menos, minimizá-las.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGUIAR, G. S. *Estudo comparativo entre Brasil e Portugal sobre diferenças nas ênfases curriculares de matemática a partir da análise do Funcionamento Diferencial do Item (DIF) do PISA 2003*. 2008. Tese (Doutorado) Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- ANDRIOLA, W. B. Estudo sobre o viés de itens em testes de rendimento: uma retrospectiva. *Estudos em Avaliação Educacional*, v. 17, n. 35, set./dez. 2006.
- _____. Principais métodos para detectar o funcionamento diferencial do item no âmbito da avaliação educacional. *Educação e Debate*, v. 2, n. 44, p.83-97, 2002.
- ANGOFF, W. H.; FORD, S. F. Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement*, v. 10, p. 95-106, 1973.
- BRASIL. Ministério da Educação. *Parâmetros Curriculares Nacionais Ensino Médio: ciências da natureza, matemática e suas tecnologias*. Brasília: MEC/Inep, 1999.
- COLE, N. S. History and development of DIF. In: HOLLAND, P. W.; WAINER, H. (Ed.). *Differential item functioning*. Hillsdale: Lawrence Erlbaum, 1993, p.
- DOOLITTLE, A. E. Gender differences in performance on mathematics achievement items. In: HOLLAND, P. W.; WAINER, H. (Ed.). *Differential Item Functioning*. Hillsdale: Lawrence Erlbaum, 1993.
- DOOLITTLE, A. E.; CLEARY, T. A. Gender-based differential item performance in mathematics achievement items. *Journal of Educational Measurement*, v. 24, p. 157-166, 1987.
- FERRER, A. T. Que variáveis explicam os melhores resultados nos estudos internacionais? In: AZEVEDO, J. *Avaliação dos resultados escolares*. Porto: ASA, 2003.
- HOLLAND, P. W.; THAYER, D. T. Differential item performance and the Mantel-Haenszel procedure. In: HOLLAND, P. W.; WAINER, H.

- (Ed.). *Differential item functioning*. Hillsdale: Lawrence Erlbaum, 1993, p. 129-145.
- IRONSON, G. H. Use of chi-square and latent trait approaches for detecting item bias. In: HOLLAND, P. W.; WAINER, H. (Ed.). *Differential item functioning*. Hillsdale: Lawrence Erlbaum, 1993.
- LINN, R. L. et al. Item bias in a test of reading comprehension. *Applied Psychological Measurement*, v. 5, p. 159-173, 1981.
- LINN, R. L.; DRASGOW, F. Implications of the golden rule settiemernt for test construction. In: HOLLAND, P. W.; WAINER, H. (Ed.). *Differential item functioning*. Hillsdale: Lawrence Erlbaum, 1993.
- LORD, F. N.; NOVICK, M. R. *Statistical theories of mental scores*. Reading, MA: Addison-Wesley, 1968.
- ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. *PISA 2003*: technical report. Paris: OECD, 2005.
- SHEPARD, L. A.; CAMILLI, G.; AVERILL, M. Comparison of procedures for detecting test item bias with both internal and external ability criteria. In: HOLLAND, P. W.; WAINER, H. (Ed.). *Differential item functioning*. Hillsdale: Lawrence Erlbaum, 1993.
- SOARES, T. M.; BAMBIRRA, F. G.; GAMERMAN, D. An Integrated bayesian model for DIF analysis. *Journal of Educational and Behavioral Statistics*, v. 34, n. 3, p. 348-377, set. 2009.
- 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. 81-102, jul./dez. 2005.
- SWAMINATHAN, H.; ROGERS, H. J. Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, v. 27, p. 361-370, 1990.
- VALLE, Raquel da Cunha. Comportamento diferencial do item: uma apresentação. *Estudos em Avaliação Educacional*, n. 25, p. 167-183, jan./dez. 2002.
- ZIMOWSKI, M. F. et al. *BILOG-MG 3 for Windows*. Scientific Software International, Chicago, 2006.

Recebido em: setembro 2009

Aprovado para publicação em: janeiro 2010

Apêndice – Itens que apresentaram DIF entre Brasil e Portugal

Código	Subárea	Competência	Contexto	Recurso gráfico	Tipo	Enunciado	Mais Fácil P/
M179Q01T	Incerteza	Conexões	Público/Social	Sim - gráfico	Aberta	Curto	Brasil
M302Q03	Mudança e Relação	Conexões	Educ./ocupacional	Sim - gráfico	Aberta	Médio	Brasil
M413Q03T	Quantidade	Reflexão	Público/Social	Não	Aberta	Médio	Brasil
M438Q02	Incerteza	Conexões	Público/Social	Sim - gráfico	Fechada	Curto	Brasil
M442Q02	Quantidade	Conexões	Público/Social	Sim - desenho	Fechada	Longo	Brasil
M484Q01T	Quantidade	Conexões	Educ./ocupacional	Sim - desenho	Fechada	Médio	Brasil
M496Q01T	Quantidade	Reflexão	Pessoal	Não	Fechada	Curto	Brasil
M520Q02	Quantidade	Reprodução	Pessoal	Sim - tabela	Fechada	Longo	Brasil
M598Q01	Espaço e Forma	Reflexão	Pessoal	Sim - figura	Fechada	Longo	Brasil
M603Q01T	Quantidade	Reflexão	Pessoal	Não	Fechada	Longo	Brasil
M806Q01T	Quantidade	Reprodução	Educ./ocupacional	Sim - desenho	Fechada	Curto	Brasil
M150Q03T	Mudança e Relação	Conexões	Científico	Sim - gráfico	Aberta	Curto	Portugal
M155Q01	Mudança e Relação	Reprodução	Científico	Sim - gráfico	Fechada	Médio	Portugal
M155Q02T	Mudança e Relação	Reprodução	Científico	Sim - gráfico	Aberta	Médio	Portugal
M273Q01T	Espaço e Forma	Conexões	Educacional/ocupacional	Sim - desenho	Fechada	Médio	Portugal
M302Q01T	Mudança e Relação	Reflexão	Educacional/ocupacional	Sim - gráfico	Fechada	Médio	Portugal
M411Q01	Quantidade	Reflexão	Educacional/ocupacional	Não	Fechada	Médio	Portugal
M413Q01	Quantidade	Reprodução	Público/Social	Não	Fechada	Médio	Portugal
M413Q02	Quantidade	Reprodução	Público/Social	Não	Fechada	Médio	Portugal
M421Q01	Incerteza	Reflexão	Educacional/ocupacional	Não	Aberta	Curto	Portugal
M438Q01	Incerteza	Reprodução	Público/Social	Sim - gráfico	Fechada	Curto	Portugal
M505Q01	Incerteza	Reflexão	Científico	Sim - tabela	Aberta	Médio	Portugal
M547Q01T	Espaço e Forma	Reprodução	Educacional/ocupacional	Sim - desenho	Fechada	Curto	Portugal