

AVALIAÇÕES EDUCACIONAIS: O PISA E O ENSINO DE CIÊNCIAS

EDUCATIONAL ASSESSMENT: PISA AND SCIENCE TEACHING

Marta F. Barroso¹, Creso Franco²

1 UFRJ/Instituto de Física/marta@if.ufrj.br

2 PUC-RJ/Departamento de Educação/creso@edu.puc-rio.br

Resumo

A avaliação internacional de estudantes realizada pela OECD, o PISA, fornece informações a respeito da aprendizagem de linguagem, matemática e ciências em muitos países do mundo. A avaliação atinge estudantes de 15 anos de idade, e é realizada em ciclos de 3 anos, a cada ciclo focando em uma componente disciplinar; pretende avaliar, na área de ciências, o denominado letramento científico. Neste trabalho, apresentamos uma análise comparativa entre países dos resultados nas questões de ciências do PISA 2000, no qual o foco foi linguagem, utilizando a teoria da resposta ao item e a identificação de questões que apresentam comportamento diferencial do item (DIF). O objetivo é verificar se o desempenho dos estudantes brasileiros têm características diferentes do desempenho de estudantes de outros países, e se estas características podem revelar diferentes ênfases curriculares no ensino de ciências. Os resultados obtidos indicam que há itens que apresentam DIF, mas não permitiram a explicação deste comportamento com base nos parâmetros escolhidos associados às ênfases curriculares.

Palavras-chave: avaliações educacionais. PISA. ensino de ciências. comportamento diferencial do item. teoria da resposta ao item

Abstract

The international student's assessment program PISA coordinated by the OECD, provides information related to language, mathematics and science learning in many countries of the world. The program has a 3 year cycle, each cycle with a focus in one of its components, and the test is applied to 15 year old students. The itens intend to evaluate the science literacy. In this paper, we present a comparative analysis of the science itens of PISA 2000, in which the focus was language literacy, using the item response theory and the identification of differential item functioning (DIF). The purpose is to evaluate if the performance of brazilian's students has different characteristics compared to those of other countries, and if these characteristics can reveal different curricular emphasis in science education. The results indicate that some itens present DIF, but we don't have an explanation of this behavior based on the choosen parameters related to the curricular characteristics.

Keywords: educational assessment; PISA; science education; differential item functioning. item response theory

Introdução

O PISA (Programme for International Student Assessment) é um programa internacional de avaliação educacional organizado pela OCDE (Organização para Cooperação e Desenvolvimento Econômico). Propõe-se a fazer uma avaliação do desempenho dos estudantes de 15 anos de idade para obter indicadores relativos aos sistemas educacionais dos países participantes. O Brasil participa, de forma voluntária, através do INEP.

Nos últimos anos, cresceu muito o número e o tipo de avaliações de aprendizagem em larga escala, especialmente de sistemas educacionais, usadas com múltiplos objetivos: comparações internacionais, avaliação somativa, comparações internas aos países, e outras [Britton e Schneider, 2006]. O tema é extremamente polêmico, por envolver questões políticas e econômicas que basicamente pressupõem tomadas de posição na maior parte das vezes baseadas em critérios ideológicos, e por prestarem-se ao uso, tanto por parte dos organismos que as organizam quanto por parte da mídia, com conotações políticas¹. Argumentos podem ser usados para os dois lados – a favor e contra a participação em, por exemplo, o exame internacional como o PISA: em um mundo globalizado, avaliações comparativas são importantes versus avaliações comparativas não ajudam e, pior, gastam-se recursos escassos; aprende-se com experiências, especialmente com as internacionais, mas as avaliações internacionais não focalizam o que é importante no plano nacional (ou regional, ou...); avaliações internacionais de grande escala contribuem para que a sociedade visualize eventuais potencialidades e problemas em seus sistemas educativos, mas as avaliações expõem muito os agentes que atuam na educação.

Para que a participação nestes exames contribua para a discussão dos sistemas educacionais no país, é necessário que eles sejam discutidos pela sociedade, que educadores e pesquisadores compreendam claramente como estes dados são obtidos e que tipo de informação pode ser obtida com a sua análise.

O que este trabalho se propõe é fornecer uma resposta, mesmo que preliminar, a algumas perguntas relativas ao ensino de ciências a partir da análise dos dados do PISA:

- Há diferenças em desempenho (avaliadas pelos objetivos do PISA) em ciências entre os alunos brasileiros e os alunos de outros países?
- Em que tipo de questões estas diferenças surgem?
- É possível associar estas diferenças a ênfases curriculares diferentes nos países?

Estas perguntas podem ser abordadas através de uso de algumas técnicas estatísticas associadas à Teoria da Resposta ao Item (ITR) [Hambleton, 1993; Hambleton, Swaminathan e Rogers, 1991], combinadas com técnicas de detecção de comportamento diferencial do item (DIF) [Soares, 2007].

As questões do PISA 2000 de Ciências foram analisadas, e concluímos que há, em alguns itens, uma diferença de desempenho entre os estudantes brasileiros e estudantes de outros países, mas não foi possível, até o momento, identificar

¹ Em relação a este aspecto, basta a recordação da recente (dezembro de 2007) divulgação no Brasil, por parte do INEP, dos resultados do PISA 2006 na área de Ciências, divididos por estados da federação.

explicações baseadas em possíveis diferenças em ênfases curriculares no Brasil. O exame de 2000 não tinha como foco o tema de ciências; é possível que a análise dos resultados do exame de 2006 permita esta identificação.

O PISA

O PISA é o Programa Internacional de Avaliação de Estudantes, organizado pela OCDE (Organização para Cooperação e Desenvolvimento Econômico). Foi organizado em ciclos, a cada 3 anos. As avaliações já feitas ocorreram nos diversos países da Comunidade Européia e em países convidados, como o Brasil, nos anos de 2000, 2003 e 2006. São avaliadas as capacidades de analisar, raciocinar e refletir sobre conhecimentos e experiências em três áreas: Leitura, Matemática e Ciências. A cada uma de suas edições, é dado foco a uma das áreas. O foco em 2000 foi Leitura; em 2003, Matemática. Em 2006, a ênfase foi para a área de Ciências. Todos os materiais (volumes, publicações e bancos de dados) são públicos e podem ser consultados na página <http://www.pisa.oecd.org>.

Os testes são, tipicamente, aplicados para um número entre 4500 e 10000 estudantes em cada um dos países participantes, escolhidos por amostragem. No Brasil, a amostragem é definida com base no Censo Escolar.

O programa busca avaliar até que ponto conhecimentos e habilidades necessárias para participar efetivamente da sociedade atual foram adquiridos pelos estudantes de 15 anos, em idade próxima ao término da escolarização obrigatória [OECD, 1999, 2000, 2001, 2003]. Define-se a amplitude dos conhecimentos, habilidades e competências a serem avaliados através do denominado "letramento" na área. Na área de leitura, foco do PISA 2000, *"Reading literacy is understanding, using, and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society."*²

Na área de Matemática, *"Mathematical literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen."*³

Na área de Ciências, *"Scientific literacy is the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity."*

Ou seja, "o letramento científico é a capacidade de usar o conhecimento científico, de identificar questões e chegar a conclusões baseadas em evidências para entender e ajudar a tomar decisões a respeito do mundo e as mudanças causadas a ele pela atividade humana."

² "Letramento em Leitura é compreender, usar e refletir a respeito de textos escritos, para atingir o objetivo individual, para desenvolver o conhecimento e potencial individual, e para participar da sociedade."

³ "Letramento em Matemática é a capacidade de um indivíduo de identificar e compreender o papel que a Matemática desempenha no mundo, de fazer julgamentos bem fundamentados e usar e se relacionar com a matemática de maneiras que atinjam as necessidades da vida individual de um cidadão construtivo, preocupado e reflexivo."

O conceito de “letramento científico”, ou “alfabetização científica”, é discutido na literatura [Roberts, 2006; Millar e Osborne, 1998], e certamente a definição adotada pela OECD não é a única.

A definição utilizada de “letramento científico” compreende três aspectos: conteúdos ou conhecimentos científicos, processos científicos e situações ou contextos nos quais os conhecimentos e processos são avaliados. Esses aspectos devem ser avaliados de forma combinada.

Para a avaliação, na área de conteúdos é feita uma seleção dos campos da ciência física, química, biológica, da Terra e do espaço de acordo com os critérios: importância para situações cotidianas; o conhecimento e as áreas de aplicação sejam de relevância duradoura para (pelo menos) a próxima década; e o conhecimento exigido possa ser combinado com os processos científicos selecionados. No aspecto de processos, que são estreitamente relacionados aos conteúdos, observam-se as habilidades e compreensão necessárias para coletar e interpretar evidências do mundo ao nosso redor, e principalmente da obtenção de conclusões, através do uso de conhecimentos científicos, após esta coleta e interpretação. No aspecto de situações ou contexto, verifica-se a capacidade de aplicação do conhecimento científico selecionado e o uso de processos científicos em situações importantes refletindo o mundo real e envolvendo as idéias da ciência.

Esta avaliação tem caráter de estudo internacional. Isto implica na necessidade de levar em conta diferenças culturais, de valores e tradições nos diversos países no qual é aplicado. Ao mesmo tempo, é necessário verificar os conhecimentos que foram obtidos no currículo escolar de ciência, principalmente, para fornecer um indicador a respeito do sistema educacional do país comparativamente aos outros.

O teste é aplicado a estudantes de 15 anos. Em países com grande defasagem série-idade, como é o caso do Brasil, muitos alunos que respondem ao teste não estão na série apropriada para a sua idade. Em muitas situações, é necessário fazer correções estatísticas (para impedir que estudantes abaixo, por exemplo, da 7ª série do ensino fundamental, ou 8º ano, sejam submetidos à avaliação). Detalhes podem ser encontrados nos relatórios do INEP [INEP 2001].

O PISA 2000 - Ciências

A prova do PISA é feita com base na Teoria da Resposta ao Item [Hambleton, 1993; Hambleton, Swaminathan e Rogers, 1991]. A prova é montada em forma de cadernos de questões, de forma que cada aluno faz um conjunto pequeno de questões diferentes.

Na avaliação de 2000, o foco era Linguagem. Isso significa que a maior parte das questões correspondia a Linguagem, e um número menor a Matemática e Ciências. A composição da nota é feita de forma a normalizar o resultado com valores entre 0 e 1000, média 500 e desvio padrão 1000 [OECD, 2004].

A prova é dividida em unidades: um texto comum (em geral, um texto de jornal ou revista, ou similar) e várias questões (itens) relacionadas. Algumas das questões são liberadas a cada ano. No ano 2000, houve 34 itens de Ciências (em 2006, o número de itens chegou a 140).

No Quadro 1, apresentamos uma unidade do exame PISA 2000 relacionada a conceitos de astronomia (Claridade). Há outras três unidades liberadas em 2003: Semmelweiss, Clonagem e Ozônio.

Quadro 1 – Uma unidade do PISA 2000

Pisa 2000 / Ciências

Exemplo de unidade do PISA 2000 - DAYLIGHT

CLARIDADE

Leia as informações abaixo e responda às questões que se seguem.

DURAÇÃO DO DIA EM 22 DE JUNHO DE 1998

Hoje, enquanto o Hemisfério Norte celebra seu dia mais longo, os australianos viverão o seu dia mais curto.

Em Melbourne*, na Austrália, o sol nascerá às 7h36 e se porá às 17h08, totalizando nove horas e 32 minutos de luz do dia.

Compare o dia de hoje com o dia mais longo do ano no Hemisfério Sul, esperado para 22 de dezembro, quando o sol nascerá às 5h55 e se porá às 20h42, totalizando 14 horas e 47 minutos de luz do dia.

O presidente da Sociedade de Astronomia, Sr. Perry Vlahos, disse que a existência de diferentes estações do ano entre os hemisférios norte e sul estava ligada à inclinação de 23 graus da Terra.

*Melbourne é uma cidade da Austrália, localizada a uma latitude de cerca de 38 graus ao sul do Equador.

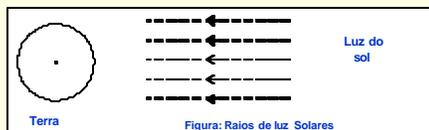
Questão 1: LUZ DO DIA S129Q1

Qual é a afirmação que explica a existência do dia e da noite na Terra?

- (A) A Terra gira em torno do seu eixo.
- (B) O Sol gira em torno do seu eixo.
- (C) O eixo da Terra é inclinado.
- (D) A Terra gira em torno do Sol.

Questão 2: LUZ DO DIA S129Q2

A figura mostra os raios de luz do sol incidindo sobre a Terra.



Suponha que seja o dia mais curto em Melbourne.

Desenhe o eixo da Terra, o Hemisfério Norte, o Hemisfério Sul e o Equador na Figura. Coloque legendas indicando cada um destes elementos.

Segundo o relatório do PISA (OECD 2004), a questão 1 foi considerada difícil, e “A resposta correta é a opção A. Esta é uma questão de múltipla escolha que exige que o estudante seja capaz de relacionar a rotação da Terra em torno de seu eixo com o fenômeno do dia e noite, e distingui-lo do fenômeno das estações, que surgem a partir da inclinação do eixo da Terra em sua rotação em torno do Sol. Todas as quatro alternativas são cientificamente corretas.”

A segunda questão desta unidade era bem mais difícil. Segundo (OECD 2004), “Esta questão é uma questão de resposta aberta que exige que os estudantes criem um modelo conceitual na forma de um diagrama mostrando a relação entre a orientação do eixo de rotação da Terra e a orientação do Sol no dia

mais curto para uma cidade do hemisfério Sul. Além disso, os estudantes devem incluir neste diagrama a posição do Equador, fazendo um ângulo de 90° com o eixo inclinado. O total de pontos (2) é obtido se o estudante coloca e indica corretamente todos os três elementos significativos: os hemisférios, o eixo inclinado e o Equador. Pontos parciais (1) são atribuídos a um diagrama com 2 dos três elementos corretamente colocados e indicados.”

Análise preliminar dos resultados do PISA 2000

Na Figura 1, apresentamos o histograma das notas do PISA 2000 para todos os países envolvidos no exame. Para a obtenção deste resultado, faz-se uma correção estatística (usando pesos) para que cada país tenha a mesma importância no resultado final, independentemente de sua população estudantil e do tamanho da amostra. Observa-se que a proficiência, normalizada com média em 500 e desvio padrão 100 para os países da OECD, varia entre os valores de 160 a 775, com média 475 e desvio padrão 101.

Na Figura 2, apresentamos o mesmo histograma, agora considerando apenas a amostra de estudantes brasileiros. Observa-se que os histogramas têm características bastante diferentes. O valor máximo em proficiência obtido por um estudante no Brasil é bastante inferior ao valor do conjunto de todos os países. A média é 379, e o desvio padrão 101.

Figura 1 – todos os países

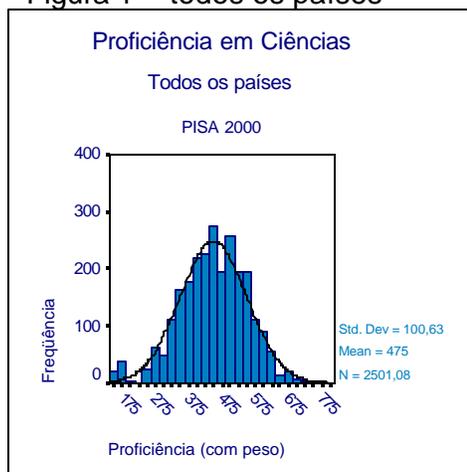
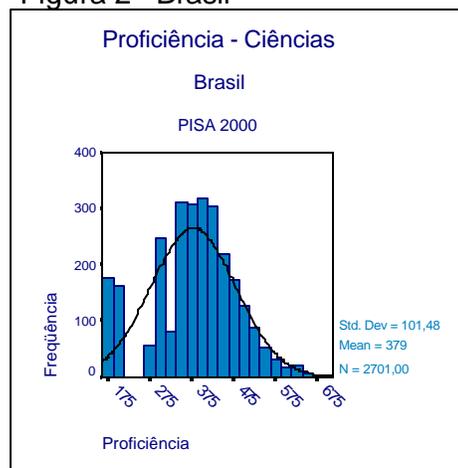


Figura 2 - Brasil

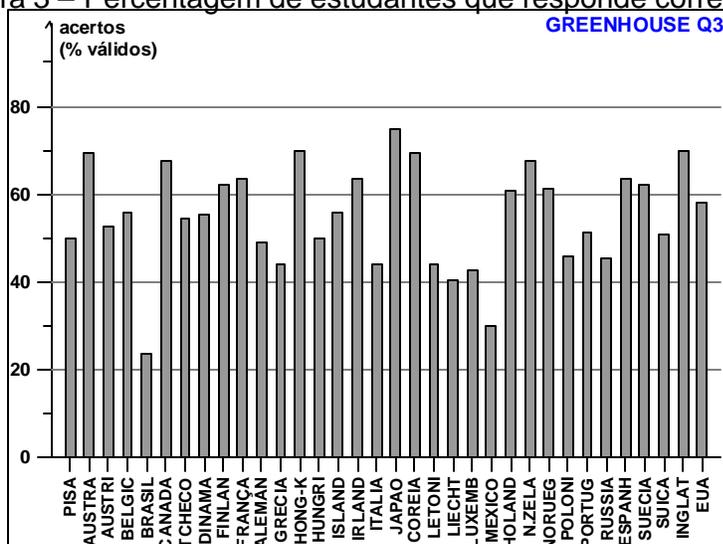


Este resultado indica o que já foi divulgado anteriormente: o desempenho global dos estudantes brasileiros de 15 anos na prova de Ciências foi bastante inferior ao resultado global. O mesmo pode se observar da análise da prova de Português [INEP, 2001] e da prova de Matemática [Aguiar, 2006]

Para entender melhor o que o PISA revela a respeito da aprendizagem de ciências no Brasil, passamos a analisar as questões. Na Figura 3, apresentamos um gráfico que revela o percentual de acertos num dado item escolhido (Efeito estufa questão 3, uma questão de resposta aberta), isto é, a percentagem do total de alunos de cada país que respondeu corretamente à questão considerada, calculado sobre o total de respostas válidas à questão.

Observa-se desta figura que os estudantes brasileiros tiveram um desempenho muito fraco, com cerca de 25% de acertos (na amostra toda, foi de 50%).

Figura 3 – Percentagem de estudantes que responde corretamente a um item



Os gráficos correspondentes para todos os outros itens apresentam um resultado similar.

As curvas características dos itens e os índices de dificuldade

Esta análise, baseada em conceitos de estatística descritiva, não nos permite nenhuma conclusão a não ser dizer que os estudantes brasileiros têm um desempenho fraco em ciências. Torna-se necessário utilizar métodos mais elaborados para conseguir entender melhor o que esses dados revelam.

A comparação de resultados de testes educacionais, entendidos como os escores que medem a proficiência dos alunos, entre grupos que apresentam culturas, características e tipos de sistemas escolares diferentes, é possível através da utilização da Teoria da Resposta ao Item nos testes aplicados [Hambleton, 1993]. Esta teoria propõe que o desempenho do estudante pode ser previsto (ou explicado) definindo-se características, que não podem ser diretamente observadas, sendo apenas mensuradas – os traços latentes. O modelo especifica uma relação entre o desempenho no item observado (resposta correta ou incorreta) e os traços ou habilidades inobserváveis que supõe-se que definam o desempenho no teste. Esta relação é chamada a curva característica do item.

A curva característica do item especifica a relação matemática entre uma habilidade e a probabilidade de acerto de um item. Como se espera que estudantes com desempenho fraco tenham pequena probabilidade de acertar um item, e estudantes com bom desempenho tenham grande probabilidade de acertar um item, esta função usualmente é parametrizada como uma curva da forma

$$P_i(\theta) = \frac{1}{1 + \exp[-y(\theta)]} = \frac{\exp[y(\theta)]}{\exp[y(\theta)] + 1}$$

onde P é a probabilidade de acerto de um item i dicotômico (isto é, para o qual só se admite a possibilidade de acerto ou erro) para um estudante com habilidade cognitiva mensurada θ . Essa curva característica é construída considerando-se, entre todos os estudantes numa determinada faixa de desempenho em ciências (por exemplo, entre 300 e 350), qual a probabilidade de acerto de um item considerado.

Isso significa que estamos fazendo comparações, num determinado item, entre alunos com a mesma habilidade cognitiva segundo a definição do teste.

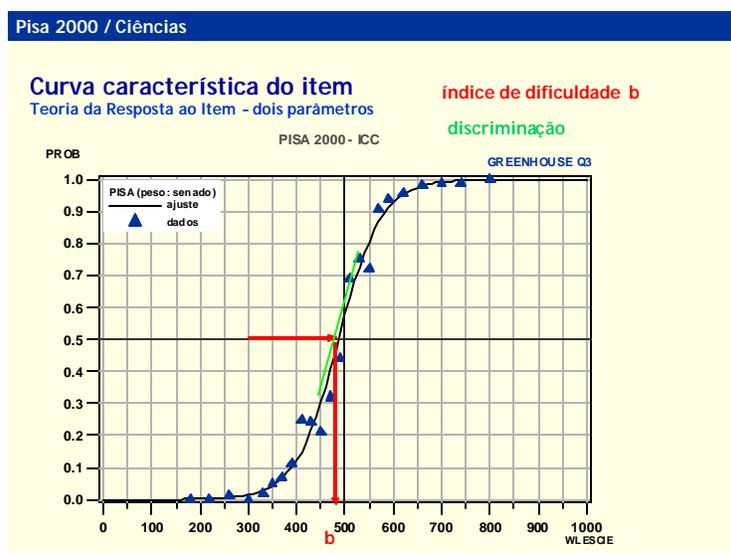
A obtenção da escala de notas do PISA é feita considerando-se a habilidade cognitiva desconhecida, e modelando $y(q)$ com um parâmetro. O ajuste fornece então o desempenho em ciências (WLESCIE) e o índice de dificuldade da questão.

Consideramos conhecida, em nossa avaliação, a habilidade cognitiva do estudante – o valor do desempenho no PISA de ciências. Foi feito então o ajuste pelo modelo de duas variáveis para $y(q)$, tratada como sendo uma função linear da habilidade cognitiva. Essas duas variáveis representam o índice de dificuldade da questão e a discriminação desta questão.

O índice de dificuldade da questão representa o valor da habilidade cognitiva para a qual o estudante tem 50% de probabilidade de acertar a questão. Se uma questão é muito fácil, ela deve ter índice de dificuldade inferior a 500.

Na Figura 4, mostramos a curva característica do item Efeito estufa Q3, o mesmo da Figura 3, para todos os países (ponderados para cada um deles ter a mesma influência, como um “senado”).

Figura 4 – Curva característica do item – Greenhouse Q3



Marcados com triângulos estão indicados os dados “empíricos”, obtidos da construção da curva, e a linha é o resultado do ajuste: $y(q) \cong b_0 + b_1 q$. Os coeficientes b_0 e b_1 são determinados por ajuste (regressão logística). O índice de dificuldade do item b_i é definido como a habilidade cognitiva para a qual a probabilidade de acerto do item vale 0.5:

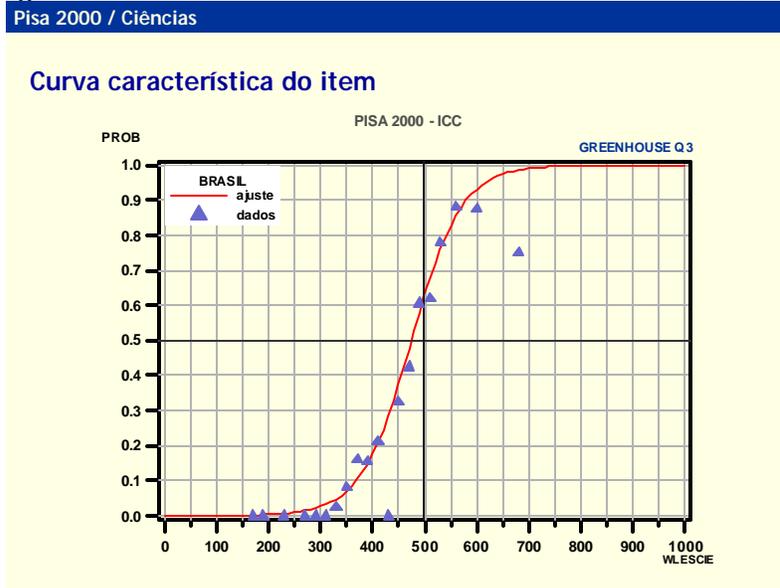
$$P_i(q = b_i) = 0.5$$

Para a função que descreve a probabilidade de acerto descrita acima,

$$y(q = b_i) = 0, \text{ ou seja, } b_i = -\frac{b_0}{b_1}$$

Na Figura 5, indicamos a mesma curva característica do item Efeito estufa Q3, apenas para os estudantes brasileiros.

Figura 5 – Curva característica do item – Greenhouse Q3



Observamos desta figura algumas características que se repetem em outros itens: há poucos estudantes brasileiros com altas habilidades cognitivas, o que transforma a parte de alta habilidade da curva (valores superiores a 650 para WLESCIE) extremamente instável por flutuações estatísticas. Também o índice de dificuldade neste item é um pouco inferior ao resultado para toda a amostra.

No entanto, esta questão não apresenta um índice de dificuldade consideravelmente diferente para os vários países. Na Figura 6, apresentamos um gráfico dos valores dos índices de dificuldade para cada país no item Efeito estufa Q3. Podemos observar que o índice de dificuldade obtido tem pequenas variações apenas, oscilando entre 450 e 510 pontos, e a média do PISA está em 480 pontos.

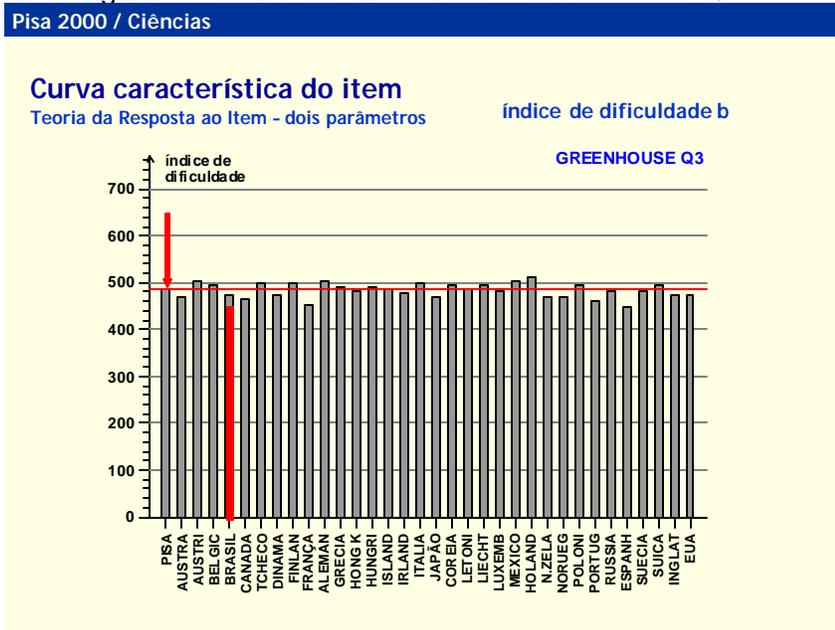
Comportamento diferencial de um item (DIF)

Para assegurar comparabilidade de resultados em exames aplicados a grupos de diferentes origens, é essencial que o modelo utilizado na avaliação educacional garanta o pressuposto de que o item apresente o mesmo comportamento nos diversos grupos populacionais que estão sendo avaliados. Para isso, pode-se fazer a análise do comportamento diferencial (DIF).

Um item apresenta DIF entre dois ou mais grupos distintos quando as probabilidades de acerto do item forem significativamente diferentes em grupos com a mesma habilidade cognitiva. A detecção de itens com DIF, para que possa ser eliminado ou diminuído seu efeito sobre a medida de proficiência (desempenho) nos processos de avaliação, é portanto bastante importante.

De forma geral, testes bem feitos não devem apresentar DIF (comportamentos diferentes em função de gênero, características étnicas, nível sócio-econômico, etc). Mas a existência de DIF num teste pode trazer informações relevantes. Por exemplo, alguns itens num teste de matemática podem ser mais difíceis para um grupo de alunos porque eles apresentam dificuldades de leitura dos textos associados aos itens.

Figura 6 – Índice de dificuldade – Greenhouse Q3

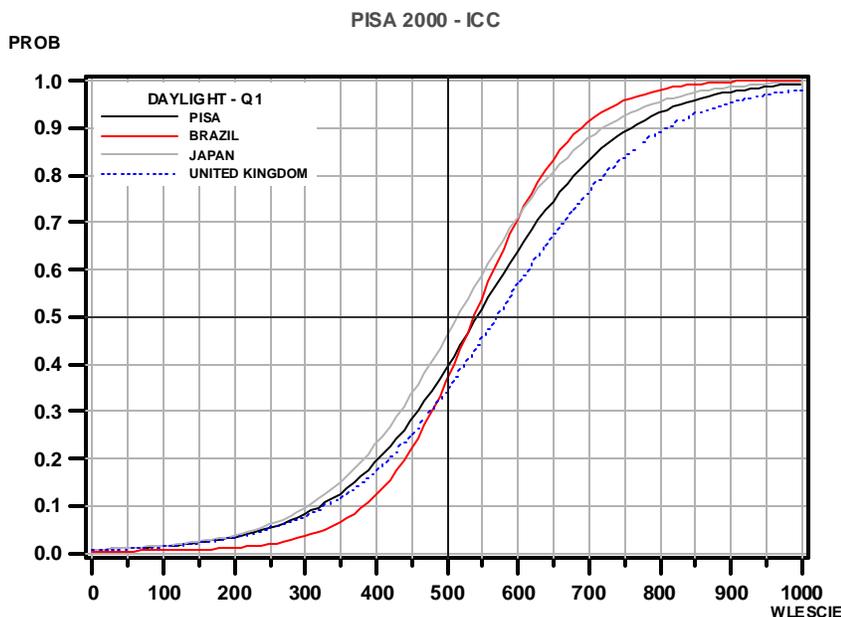


A explicação da existência de DIF é muitas vezes difícil, “mas é consenso que o DIF, quando elevado, pode trazer implicações técnico-estatísticas e, às vezes, embaraços sociais ou de cunho moral. Não obstante esses problemas, os resultados encontrados podem representar uma oportunidade de estudos extremamente relevantes que possibilitem, entre outros fatores, avaliar diferenças curriculares, de abordagens pedagógicas e diferenças étnico-sócio-culturais.”[Aguiar, 2006].

A detecção dos itens com DIF, e com a conseqüente eliminação ou diminuição de seus efeitos sobre a medida de proficiência, é a justificativa para muitos métodos propostos [Soares, 2007]. Mas muitas vezes a existência de DIF revela informações importantes. A análise dessa existência, com a identificação dos fatores associados ou mesmo a identificação de sua magnitude, permite estudos sobre diferenças regionais sócio-culturais, ênfases curriculares diversas, diferentes abordagens pedagógicas, entre outras, que não são facilmente percebidas.

A análise dos itens do PISA 2000–Ciências permite identificar um conjunto de itens que apresenta DIF. Por exemplo, na Figura 7 apresentamos a curva característica do item Claridade Q1 (ver texto no Quadro 1). A partir desta figura, podemos observar que tanto o índice de dificuldade (o valor da habilidade cognitiva, determinado pela variável WLESCIE, para o qual a probabilidade de acerto do item é 0,5) quanto a discriminação do item (a inclinação da curva característica do item no ponto para o qual a probabilidade de acerto é 0,5, e que indica a variação de habilidade cognitiva para que haja uma diferença significativa entre a probabilidade de acertar ou de errar o item) apresentam DIF na comparação dos resultados globais (PISA, na figura), do Brasil, do Japão e do Reino Unido.

Figura 7 – Curva característica do item – questão Claridade Q1 – para alguns países



Conclusões e desenvolvimentos

A análise feita, baseada na existência de DIF e na teoria da resposta ao item, permitem afirmar que o PISA 2000 – Ciências revela diferenças de desempenho entre estudantes brasileiros e estudantes de outros países. Há um conjunto de itens em que estas diferenças se manifestam, se escolhermos a detecção de DIF como a técnica para definir a existência destas diferenças.

Um dos objetivos do trabalho, identificar as ênfases curriculares que poderiam explicar essas diferenças, porém, ainda não foi atingido. O número de itens do PISA 2000 foi pequeno (34).

Buscamos explicar o DIF com base em aspectos relacionados à apresentação do item (existência de imagens, tamanho do texto, gráficos), aos conteúdos (física, biologia, química, geociências), aos processos avaliados (identificar evidências, comunicar ou avaliar conclusões), às situações nas quais o item é apresentado (terra e ambiente, vida e saúde, ciência e tecnologia), ao contexto (individual, coletivo ou global), à necessidade de trabalhos experimentais ou práticos para melhor apreensão do item, e aos valores do índice de dificuldade e coeficiente de discriminação. Numa análise inicial não foi possível detectar nenhuma correlação entre esses aspectos e a existência de DIF.

Necessita-se então de técnicas estatísticas mais elaboradas [Soares 2007] e utilização dos dados do PISA 2006 para avançar nesta avaliação.

De qualquer forma, foi possível determinar alguns componentes bastante interessantes no ensino de ciências em nosso país, comparativamente aos outros. Os níveis máximos de habilidade cognitiva avaliados pelo PISA não são alcançados pelos estudantes brasileiros. Em boa parte das questões, a probabilidade dos alunos brasileiros com uma dada habilidade cognitiva acertarem um item é similar a dos alunos dos outros países. Apenas uma parcela dos itens apresenta comportamento diferencial do item, e no momento ainda não temos uma explicação bem embasada para a existência desta diferença em apenas alguns dos itens.

Referências

- AGUIAR, Glauco da Silva. Projeto de Exame de Qualificação ao Doutorado, “Comportamento Diferencial dos Itens (DIF) como Estratégia de Apreensão de Ênfases Curriculares em Matemática em Diferentes Países”, Programa de Pós-Graduação em Educação – Doutorado, PUC-RJ, 2006.
- BRITTON, Edward D.; SCHNEIDER, Steven A. Large-Scale Assessments in Science Education. In: ABELL, Sandra K. e LEDERMAN, Norman G. (ed). **Handbook of Research on Science Education**. New Jersey: Lawrence Erlbaum Associates, Publishers, 2006, p. 1007-1042.
- HAMBLETON, Ronald K. Principles and Selected Applications of Item Response Theory. In: LINN, Robert L. (ed.), **Educational Measurement**, 3rd edition, Phoenix: American Council on Education, Orix Press, 1993, p. 147-200.
- HAMBLETON, Ronald; SWAMINATHAN, H.; JANE ROGERS, H. **Fundamentals of Item Response Theory**. Newbury Park: SAGE Publications, 1991.
- INSTITUTO NACIONAL DE ESTATÍSTICAS EDUCACIONAIS – INEP. **Pisa 2000: Relatório Nacional**. 2001. Disponível em < <http://www.inep.gov.br/download/internacional/pisa/PISA2000.pdf> >. Consultado em 22 de abril de 2008.
- MILLAR, Robin; OSBORNE, Jonathan (ed.). **Beyond 2000: Science Education for the Future** (The report of a seminar series funded by the Nuffield Foundation), 1998. Disponível em < <http://www.kcl.ac.uk/content/1/c6/01/32/03/b2000.pdf> >. Acesso em 27 de abril de 2008.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Knowledge and skills for life: first results from PISA 2000**. Paris: OECD, 2001.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Learning for tomorrow’s world – first results from PISA 2003**. Paris: OECD, 2004
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Measuring students knowledge and skills: the Pisa 2000 assessment of reading, mathematical and scientific literacy**. Paris: OECD, 2000.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Measuring students knowledge and skills: a new framework for assessment**. Paris: OECD, 1999.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills** - Publications 2003. Paris: OECD, 2003. Disponível em <<http://www.oecd.org/dataoecd/46/14/33694881.pdf>>. Consultado em 30/10/2006.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **PISA 2003 Data Analysis Manual – SPSS Users**. Paris: OECD, 2004. Disponível em <http://www.pisa.oecd.org/document/18/0,3343,en_32252351_32236173_35016146_1_1_1_1,00.html >. Consultado em 24 de abril de 2008..
- ROBERTS, Douglas. Scientific Literacy/Science Literacy. In: ABELL, Sandra K. e LEDERMAN, Norman G. (ed). **Handbook of Research on Science Education**. New Jersey: Lawrence Erlbaum Associates, Publishers, 2006, p. 729-780.
- SOARES, Tufi Machado. O funcionamento diferencial do item e suas implicações educacionais. Comunicação privada, 2007.