Introdução

As últimas décadas do século XX foram marcadas por uma evidente preocupação com a melhoria da qualidade da educação. Cresceram e se consolidaram formas de cooperação internacional, cujos organismos mais influentes (UNESCO, OCDE, Banco Mundial) configuraram-se em espaços essenciais à formulação de políticas educacionais em diferentes países. No Brasil, essa preocupação com a educação culminou com a implementação dos sistemas de avaliações educacionais que têm como um dos principais objetivos acompanhar o comportamento do sistema educacional, no que diz respeito à eficácia e à equidade dos resultados educacionais. Tais empreendimentos, dentre outros aspectos, estão relacionados com a comparação do desempenho dos alunos em diferentes séries e disciplinas, em momentos distintos.

Neste contexto de renovada preocupação com a educação, encarada, mais uma vez, como fator imprescindível na busca pelo desenvolvimento econômico e bem-estar social das populações, face às grandes transformações ocorridas nas relações econômicas e sociais internacionais, o Brasil amplia seus interesses na área de avaliação educacional. Com o intuito de comparar os alunos brasileiros com os de outros países, voluntariamente, passa a participar, a partir de 2000, do PISA – *Programme for International Student Assessment*. Trata-se de uma pesquisa internacional de avaliação de alunos de 15 anos de idade, realizada pela OCDE - Organização para Cooperação e Desenvolvimento Econômico. Esse estudo tem como principal objetivo avaliar o nível de competências cognitivas dos estudantes de diversos países a partir de um teste nas áreas de Leitura, Ciências e Matemáticas. Outro objetivo desta avaliação é verificar a quanto os sistemas educacionais dos países participantes estão cumprindo a função de democratização e acesso ao conhecimento, mensurando também o grau de desigualdade existente entre eles. Em suma, o PISA tem como objetivos principais: elaborar um sistema internacional de indicadores da educação e comparar os resultados e a situação educacional dos diferentes países. (OCDE,2005).

O PISA não é o primeiro estudo internacional sobre desempenho cognitivo de estudantes, em uma perspectiva comparada. Seus antecessores remontam ao final dos anos cinquenta, quando foi criada a *International Association for the Evaluation of Educational Achievement (IEA)* por um grupo de notáveis pesquisadores em Educação. A proposta destes consistia em realizar estudos internacionais sobre o rendimento dos estudantes, concentrando a avaliação nas partes do currículo essencialmente comuns aos países participantes. Por isso, aspectos de conteúdos específicos do currículo de algum país, ou de um número pequeno de países, normalmente não eram considerados nos testes, a despeito de o quanto significantes essas partes do currículo fossem para esses poucos países. Para além de calcular e avaliar os resultados, havia também um esforço para que se identificassem os principais fatores associados às diferenças de rendimento e para se analisar o impacto desses fatores, a fim de se subsidiarem políticas de melhorias na educação (Ferrer, 2003).

A metodologia adotada pela IEA, basicamente quantitativa, serviu de referência e modelo para desencadear uma série de novos estudos desta natureza, tanto em escala mundial como circunscritos a algumas áreas geográficas ou geopolíticas concretas, como a União Européia e o Mercosul. É marcadamente nos anos oitenta, incrementada pelos organismos internacionais, que elegem a educação como principal fator gerador do crescimento econômico e do bem-estar social, que essa febre avaliativa ganha fôlego novo. Conseqüência do compromisso assumido com a melhoria da qualidade da educação, cujas ações não podem prescindir da avaliação de desempenho e das variáveis chaves para as explicações ou interpretações dos resultados.

Esses estudos internacionais mais recentes têm características diferentes, não só pelo formato, periodicidade ou população envolvida, mas principalmente pelo que avaliam, exatamente. No entanto, de um modo geral, englobam dois objetivos fundamentais que não são necessariamente antagônicos ou excludentes. Por um lado, buscam a comparação dos diversos sistemas educacionais, no que se refere a sua estrutura ou funcionamento. Por outro lado, objetivam a explicação e interpretação dos resultados alcançados. Quando o foco está no primeiro objetivo, esses estudos se caracterizam pela construção de indicadores da área educacional. Quando se prioriza o segundo objetivo, os estudos buscam a identificação e interpretação dos fatores que impactam o desempenho dos estudantes,

normalmente numa combinação com outros dados coletados. O PISA, colocado em prática pela OCDE, no final dos anos noventa, é um estudo que combina esses dois objetivos. Ao mesmo tempo em que torna possível, aos pesquisadores dos diferentes países participantes, as análises sobre os fatores que impactam a proficiência dos estudantes, responde às preocupações dos gestores de diferentes nações, quanto ao desenvolvimento do capital humano da população, no que tange às taxas alcançadas, níveis de instrução e rendimento.

É justamente essa busca em medir o grau de desenvolvimento do capital humano das populações dos diferentes países envolvidos no estudo, que imprime ao PISA um certo caráter de originalidade e pioneirismo. Guardadas as similaridades com o conceito de “capital humano” de Coleman e com a definição de Bourdieu para “capital cultural institucionalizado”, o conceito de capital humano adotado pelo PISA é um pouco mais complexo. Não deve ser entendido apenas como o nível de instrução das pessoas ou os títulos escolares obtidos por estas. O conhecimento, as habilidades e competências, definidos pela OCDE, e foco da avaliação, medidos nos testes cognitivos, têm mais a ver com as virtudes pessoais, competências diversas, atitudes e disposições sociomotivacionais do que com a aprendizagem escolar sistemática, assimilada de forma automática e sem conexão com o mundo real. Na concepção dos pesquisadores do PISA, o conceito, tal como proposto em meados do século XX, perdeu muito de seu sentido. Afinal, embora a educação continue sendo vista como fundamental para o desenvolvimento social e econômico, muda, na visão dos gestores, a chave para a solução dos problemas por ela suscitados. O esforço pela qualidade se desloca mais em direção à busca por uma aprendizagem eficaz, envolvendo atributos necessários a uma plena participação social e democrática na vida adulta que habilite as pessoas a se tornarem aprendizes ao longo da vida do que, propriamente, apenas alcançarem maiores índices de diplomação e medir a contribuição da educação para o crescimento econômico (OCDE, 2005).

Ao avaliar as competências e habilidades ao fim da escolarização básica, o PISA examina o grau de preparação dos jovens para a vida adulta e, até certo ponto, a efetividade dos sistemas educacionais. 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.” (PCN, p.10). Cabe ressaltar que, essa matriz conceitual de avaliação do PISA, que tem como referência principal a articulação entre o conceito de educação básica e o de cidadania, coaduna-se com os objetivos educacionais do Ensino Médio, já propostos pelo Ministério da Educação - MEC. Ela está presente também no ENEM – Exame Nacional do Ensino Médio - cujos preceitos demonstram seu caráter transdisciplinar e o entendimento do perfil terminativo desse nível de ensino, complementando o aprendizado iniciado no Ensino Fundamental. Essa estrutura de avaliação é desenhada a partir de um modelo dinâmico de aprendizagem, em que, tal como definidos nos textos constitucional e na Lei de Diretrizes e Bases - LDB de 1996, os conhecimentos e habilidades devem ser continuamente adquiridos, a fim de viabilizarem uma adaptação bem-sucedida, em um mundo em constante transformação.

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. Chamados a se posicionarem, alguns especialistas e autoridades apontam como possíveis causas para o problema a formação deficiente dos professores e a infraestrutura precária das escolas. Outros analisam esses resultados à luz das variáveis extrínsecas ao sistema, como a comparação do Índice de Desenvolvimento Humano - IDH dos países, o nível socioeconômico e cultural das populações e os recursos financeiros alocados para a educação. Fica evidente, portanto, que, a despeito dos cuidados metodológicos de que se cercam os idealizadores do PISA, existe uma dificuldade matricial em abarcar a imensa diversidade de contextos que o estudo comporta. Por outro lado, vale destacar as potencialidades do estudo, na elaboração de um sistema internacional de indicadores da educação e na comparação da situação educacional dos diferentes países, podendo este sistema vir a se constituir em uma referência útil, se utilizado para a identificação de pontos fortes e pontos fracos no monitoramento dos progressos alcançados ao longo do tempo.

Essa comparação de resultados nos testes educacionais, entendido como o resultado dos escores que medem a proficiência dos alunos, é possível graças à

utilização da Teoria de Resposta ao Item- TRI - nos testes aplicados. Nesta teoria, modelos matemáticos relacionam a probabilidade de um aluno responder a um item corretamente e suas habilidades na área do conhecimento avaliado, entendidas estas como um conjunto de traços latentes, não observados diretamente. No caso do PISA, a TRI possibilita a comparação entre alunos de diferentes países, uma vez que o modelo empregado nos itens é conhecido, esses estão na mesma escala e há itens comuns nos testes aplicados aos diferentes alunos dos países participantes.

De acordo com Soares (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 comportamento 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 (*DIF-Differential Item Functioning*). Um item apresenta DIF entre dois ou mais grupos distintos, quando esses 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. Esse é um tipo específico de DIF, associado ao grau de dificuldade do item. Assim, na estimação das proficiências, o ideal é evitar a utilização de itens com DIF elevado, isto é, os quais favoreçam um determinado grupo de alunos, em detrimento de outros. Todavia, o DIF, quando moderado e restrito a poucos itens, interfere minimamente na estimação da proficiência e sua análise pode ser uma ferramenta de diagnóstico do sistema educacional bastante útil ao que se refere às diferenças curriculares, socioculturais e, no caso de estudos internacionais, como o PISA, à diversidade de realidades educacionais e à disparidade de resultados entre países.

Em síntese, a análise de DIF tem por objetivo investigar se um item apresenta graus diferentes de dificuldade, para subgrupos que têm a mesma habilidade. Caso isso ocorra, tal fato pode significar que algum grupo particular de indivíduos estaria sendo privilegiado, em detrimento dos demais. Entretanto, a evidência estatística de uma grande diferença entre o desempenho de grupos não deve ser, ainda assim, persuasiva quanto à decisão de excluir ou não um item do

teste, pois pode haver boas razões para que se mantenha o item, apesar das estatísticas. 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 ou meninas, devem saber lidar com problemas dessa natureza. Excluir tais itens de um teste, por favorecerem um determinado grupo, torna o instrumento incompleto e traz prejuízo para os que estão em desvantagem, pois tende a perpetuar a diferença. Shepard (1987a) expressou ponto de vista similar: *“na medida em que membros de um grupo têm menos capacitação para responder a um conjunto de itens que são uma parte legítima do domínio de conteúdo, devemos ser cautelosos em afirmar que a discrepância evidencia viés”* (p. 186, em itálico no original).

Assim, da mesma forma que as ferramentas estatísticas podem evitar a parcialidade nos testes, minimizando as implicações técnico-estatísticas e, às vezes, os embaraços de cunho moral e éticos, a análise do DIF, quando este existir, pode fornecer pistas importantes em relação ao processo de aprendizagem de determinado conteúdo. De acordo com estudos do *Educational Testing Service - ETS*, nos Estados Unidos, o DIF pode ser causado, basicamente, por três fatores:

- Familiaridade do respondente com o conteúdo do item;
- Interesse pessoal do respondente por um determinado conteúdo;
- Reação emocional negativa do respondente provocada pelo conteúdo do item.

Cabe ressaltar que, a explicação para a existência de DIF em um item nem sempre é trivial; mas, quando encontrada, abre novas oportunidades de estudos, singulares, uma vez que possibilita, entre outros aspectos, avaliar diferenças curriculares, de abordagens pedagógicas e diferenças etno-socioculturais.

Embora ciente de quão árdua possa ser a tarefa de encontrar uma explicação para a existência de DIF em um determinado item, este foi o foco norteador do estudo que ora desenvolvo, tendo como base os itens de Matemática do PISA 2003. Este estudo está relacionado com o tema maior da qualidade da Educação Básica no Brasil, cujos objetivos envolvem: a) a avaliação do grau de desigualdade e iniquidade educacionais no Brasil; b) comparação de resultados dos alunos brasileiros com os de outros países, no que tange à defasagem entre as reais competências de nossos alunos e as exigidas pelos países mais avançados,

para o exercício pleno da cidadania. Dentro de uma perspectiva de que essas avaliações não devam apenas limitar-se a medir e comparar os desempenhos conseguidos pelos alunos, mas também interpretá-los, meu objetivo foi identificar os principais fatores capazes de explicar as diferenças de rendimentos encontradas no letramento em Matemática. Mais precisamente, o meu propósito foi investigar as características dos itens de teste, em relação à forma de apresentação, aos conteúdos e às áreas da Matemática no sentido de apontar ênfases curriculares diferenciadas dos países participantes do PISA, as quais pudessem dar pistas para as diferenças de rendimento entre os estudantes.

Para a realização desse estudo, os objetivos desdobraram-se nas seguintes questões de pesquisa:

1- Há diferenças de competência matemática entre os alunos brasileiros e os de outros países?

2- É possível identificar conteúdos ou competências matemáticas, os quais, possivelmente, não são igualmente explorados nos currículos dos países avaliados?

3- Quais são os itens da prova de Matemática que apresentam DIF, tendo o Brasil como grupo de referência? Quais as dimensões (magnitudes) dos DIFs encontrados?

4- Os itens que apresentam DIF seguem algum padrão? Referem-se a alguma especificidade curricular (conteúdos ou competências matemáticas) ou dizem respeito a diversidades etno-culturais?

Embora um grande esforço seja despendido pelos especialistas em avaliação, no sentido de se construírem instrumentos não afetados por diferenças culturais, sejam estas decorrentes de idiomas ou de outro tipo de ruído qualquer, os itens do PISA apresentam funcionamento diferencial (DIF) entre os diferentes grupos. Considerando que os sistemas diferem entre si, com uma variação que se pode considerar natural, e que as características próprias que os distinguem têm conseqüências nos diversos modos de elaboração e desenvolvimento do currículo; e, ainda, que os conteúdos são selecionados pelos professores e abordados com ênfases diferenciadas, o estudo em questão buscou identificar fatores que possam interferir no desempenho dos alunos, por razões que extrapolam, *per si*, a proficiência de cada um. Sendo a análise de DIF 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, e na utilização de um modelo de regressão cuja característica é a possibilidade de confirmação ou não das hipóteses levantadas para a explicação do DIF.

Esta tese encontra-se estruturada em cinco capítulos, incluindo esta introdução. No segundo capítulo, tendo Forquin como referência teórica, apresento o conceito de currículo ao qual nos deteremos. Em seguida, a partir da série histórica do SAEB de 1995 a 2005, exponho os resultados obtidos em Matemática pelos alunos brasileiros, apontando os problemas do ensino e da aprendizagem dessa disciplina, ao longo do tempo, e discutindo o papel das avaliações em larga escala como estratégia de apreensão do currículo aprendido. Além disso, procuro descrever o que é o DIF (seus conceitos, tecnalidades e contexto histórico) e cito alguns estudos clássicos na área de análise de DIF.

O terceiro capítulo, de caráter estritamente descritivo, traz uma apresentação da avaliação internacional PISA, explicitando suas características, particularidades e seus objetivos.

No quarto capítulo, fazendo uso da escala de proficiência em Matemática do PISA 2003, apresento e discuto os resultados alcançados por nossos alunos e seus colegas portugueses que participaram da avaliação. Ainda, objetivando esclarecer sobre o conceito de “letramento”, tipo de instrumento utilizado e os domínios de avaliação cobertos pelo PISA, em termos de “conteúdo”, “processos” e “contextos”, alguns itens públicos utilizados na avaliação, serão apresentados e discutidos.

O quinto capítulo é dedicado aos resultados das análises de DIF e ao estudo exploratório das características dos itens que podem explicar o funcionamento diferencial de alguns itens em relação aos diferentes grupos.

Finalizo com a retomada dos principais achados dos capítulos 4 e 5, discutidos à luz da literatura sobre Educação Matemática. Procuro mostrar até que ponto os resultados da análise de DIF estão em sintonia com as visões usuais a respeito dos desafios da Educação Matemática na escola. Além disso, indico as limitações do estudo e destaco como a análise de DIF pode ser uma útil ferramenta de diagnóstico do sistema educacional, avaliando diferenças regionais

e sócio-culturais de difícil percepção, além de diferenças curriculares e diferenças de abordagens pedagógicas.

Esclareço ainda que, pelas características deste trabalho, a discussão dos aspectos metodológicos estará presente ao longo dos diversos capítulos.