

Avaliação educacional em larga escala: a experiência brasileira

FERNANDA DA ROSA BECKER
Pesquisadora do INEP, Ministério da Educação, Brasil

1. Introdução

Ao longo da década de 90, iniciou-se um ciclo de processos de reformas baseado na ideia de que era necessário modificar o desenho organizacional e institucional dos sistemas educacionais. Um dos argumentos que sustentava esta ideia era o baixo nível de responsabilidade por resultados com que operavam as administrações tradicionais (TEDESCO, 2003).

A responsabilização na educação foi apenas mais uma face do movimento mais amplo na gestão pública. A responsabilização passa a ser entendida como um valor que deve guiar os governos democráticos uma vez que é uma forma de rendição de contas a sociedade.

Um dos primeiros modelos de *accountability*- responsabilização- na educação a serem desenvolvidos foi o modelo inglês criado em 1988 por meio do *Education Reform Act*. Houve a centralização do currículo, criação de sistemas de avaliação e aos poucos as escolas passaram a ter mais liberdade para gerir os recursos recebidos. A ideia central era fortalecer a aprendizagem sem desrespeitar a função exercida pelo professor ao mesmo tempo em que se criava um modelo de avaliação para atender à demanda por responsabilização (EARL, 1999).

Deste modo, muitos governos passaram a se preocupar em coletar e divulgar estatísticas que retratassem o funcionamento do sistema educacional como um todo, mas não realizavam uma avaliação do produto final da educação, isto é, o que tinha sido realmente aprendido pelos alunos.

Atualmente este enfoque está mudando e alguns países já realizam avaliações nacionais. Esta mudança começou após análises que indicavam que as crianças aprendiam pouco ao se considerar o tempo que passavam na escola. Outra razão para a mudança de foco das avaliações é o fato de muitos governos, caso brasileiro, depararem-se com o desafio de expandir o sistema ao mesmo tempo em que expandem a qualidade.

O presente estudo tem por objetivo apresentar a experiência brasileira na área de avaliação educacional em larga escala. Dessa forma, o artigo se divide em cinco seções das quais a primeira é esta introdução; a segunda apresenta a problemática da avaliação educacional em larga escala, a terceira introduz a experiência brasileira a nível nacional, a quarta a experiência no âmbito internacional e a última traz alguns comentários finais.

Revista Iberoamericana de Educación / Revista Ibero-americana de Educação
ISSN: 1681-5653

n.º 53/1 – 25/06/10

Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI)

Organização dos Estados Ibero-americanos para a Educação, a Ciência e a Cultura (OEI)



2. Avaliação Educacional em Larga Escala

Ferrer (2006) descreve algumas condições desejáveis para que um país tenha um sistema de avaliação. Em primeiro lugar, os propósitos das avaliações e o uso do material coletado devem ser previamente debatidos para que as provas e a metodologia a ser desenvolvida atendam estes princípios.

Em segundo lugar, os instrumentos de medição devem estar relacionados a metas e objetivos de desenvolvimento cognitivo traçados pelo país. Logo, devem existir metas que sejam conhecidas por todos: professores, diretores e sociedade em geral.

Outro ponto ressaltado no estudo é a necessidade de decidir se o objetivo é avaliar o sistema, as instituições ou os atores individualmente, para que seja escolhido o modelo de avaliação (censitária ou amostral). Ademais, todos os instrumentos devem ser validados antes da avaliação em si. Recomenda-se realizar antes um projeto "piloto".

Por fim, é feita ressalva com relação à forma de apresentar os resultados. Esta deve ser coerente com o objetivo das medições. Não adianta divulgar um dado por escola se o objetivo é identificar alunos problemáticos, por exemplo.

Um ponto importante e normalmente utilizado como argumento contrário à implementação de sistemas de avaliação é o financiamento, isto é, o custo para que um país realize estes exames (SAUS, 2003). Iaies (2003) afirma que parecem estar mais claros os custos do que os benefícios dos sistemas de avaliação e sugere uma revisão crítica uma vez que as avaliações passam a ser objeto de avaliação.

Em contrapartida, Wolff (2007) analisa o caso de Chile, Colômbia, Honduras, Peru e Uruguai. São dados de avaliações censitárias no Chile e na Colômbia; avaliações amostrais em Honduras, Uruguai, Colômbia, Peru e avaliações internacionais (amostrais) no Chile, Colômbia e Peru.

O estudo mostra que os custos variam significativamente entre os países mesmo quando se avalia aproximadamente o mesmo número de alunos uma vez que o número de matérias e séries avaliadas varia de país para país, assim como o número de informações coletadas. Alguns países coletam informações com relação ao *background* dos alunos e demais fatores que afetam o aprendizado. Isso aumenta os custos.

Há que se considerar também a diferença de extensão territorial e de distribuição da população. Além disso, alguns países não possuem uma equipe para avaliar os resultados. Então recorrem a consultores nacionais e internacionais, o que incorre em aumento dos custos.

Apesar de haver grande diferença de custo entre países, em nenhuma das análises realizadas pelo autor o valor de implementação do sistema de avaliação chegou a 0,3% do orçamento da educação do nível avaliado. Logo, avaliar está entre as inovações menos custosas da reforma educativa.

Atualmente 16 países realizam avaliações censitárias ou amostrais na América Latina. O Chile avalia regularmente de forma censitária alunos de séries selecionadas assim como México, Colômbia e Brasil. El Salvador e Guatemala iniciaram recentemente este processo. Os resultados dessas avaliações são

utilizados de forma diferente nos diversos países. No Chile e no México os resultados dos estudantes têm impacto no salário dos docentes e nas possibilidades de ascensão na carreira.

Com relação à prática destes países, Wolff (2007) afirma que é difícil concluir por meio de uma avaliação destas práticas se um professor ou uma escola é melhor do que outros. A questão-chave são os antecedentes individuais. Um professor que consegue que seu aluno, filho de analfabetos, alcance uma pontuação média, é muito mais efetivo que um professor que faz com que um aluno, filho de pais com ensino superior, atinja 20% acima da média. No entanto, isso não é medido pela nota individual.

O Sistema Nacional de Avaliação da Educação Básica (SAEB) foi a primeira iniciativa brasileira no sentido de conhecer a fundo os problemas e deficiências do sistema educacional, para orientar com maior precisão as políticas governamentais voltadas para a melhoria da qualidade do ensino.

Criado em 1990, teve seu segundo ciclo de aplicação em 1993, mas somente a partir de 1995 adquiriu um papel estratégico, articulando-se com um conjunto de políticas voltadas para a melhoria da qualidade do ensino em todos os seus níveis (UEMA, 2003). A avaliação, que era uma diretriz governamental, tornou-se clara atribuição do Ministério da Educação com a Lei de Diretrizes e Bases da Educação de 1996.

O SAEB tornou possível identificar os problemas do ensino e suas diferenças regionais por meio de dados e indicadores que possibilitam uma maior compreensão dos fatores que influenciam o desempenho dos alunos e proporcionou aos agentes educacionais e à sociedade uma visão concreta dos resultados dos processos de ensino e aprendizagem e das condições em que são desenvolvidos.

As avaliações nacionais tanto podem ser um instrumento para que as escolas, gestores e demais profissionais repensem seu trabalho, quanto para se efetuar classificações que pouco contribuem para a melhora da qualidade das escolas. Se a avaliação for concebida apenas como mecanismo de controle ou se for ignorada, seja pelos professores, pelos gestores ou pelos responsáveis pela formação de docentes, pouco contribui para a melhora do ensino (SOUSA, 2003).

Destarte, a avaliação não é um fim em si mesmo, mas um instrumento que deve ser utilizado para corrigir rumos e pensar o futuro. É crucial assegurar que, juntamente com as informações que a avaliação fornece, sejam criados e utilizados instrumentos que contribuam substancialmente para a solução dos sérios problemas sociais que afetam a população em idade escolar.

Segundo Uema (2003), é preciso fazer a ponte entre a avaliação e a ação. Não basta informar ou iniciar um processo de reflexão na escola sobre os problemas ou deficiências pelos quais ela passa, sem colocar à sua disposição os recursos de que precisa para superar suas dificuldades, pois, nesse caso, a avaliação tende a exercer somente o papel inútil de testemunha, e não de agente transformador da realidade.

Esta dificuldade em fazer a "ponte" também aparece em relatórios sobre os sistemas de avaliação educacional de outros países. Kellaghan & Greaney (2004) reportam o enorme hiato entre o que as avaliações demandam dos professores e a visão dos professores a respeito do papel deles no processo de avaliação nos países africanos. Outro ponto destacado pelos autores é a questão da vontade política uma

vez que, sem apoio do governo, possíveis recomendações que possam ser feitas, com base nos dados levantados, não serão postas em prática.

Como exposto por Ferrer (2006), devem existir metas nacionais, portanto, um compromisso assumido e conhecido por todos: professores, diretores e sociedade em geral. No caso de metas mais gerais e não um objetivo preciso seria interessante aproximar a equipe técnica de avaliação e os profissionais de educação.

Uma avaliação feita das avaliações educacionais da América Latina, ao longo da década de 90, concluiu que os docentes perceberam os mecanismos de avaliação mais como uma forma de pressão do que como um insumo para a melhora da qualidade da educação. Desta forma, os resultados não se transformaram em instrumentos de políticas públicas e o impacto sobre o ensino nas escolas foi muito baixo (TEDESCO, 2003). Os professores parecem estar de acordo com a ideia da avaliação, mas têm críticas com relação à maneira pela qual os processos foram implementados (BID, 2002; IAIES, 2003). Pode ser uma missão difícil fazer com que todos os professores conheçam as metas (espera-se que existam metas bem definidas) e desenvolvam estratégias para alcançá-las, mas é essencial que isto seja feito.

Esta seção introduziu o debate a respeito do uso de sistemas de avaliação por meio de um breve histórico e algumas críticas existentes na literatura. As próximas seções abordarão o histórico das avaliações nacionais e internacionais no Brasil.

3. Avaliações Nacionais da Educação Básica no Brasil

A primeira avaliação nacional de larga escala da educação básica no Brasil foi o SAEB. O sistema é coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e foi implantado em 1990. O exame tem caráter amostral e atualmente avalia os alunos de 4ª e 8ª séries do Ensino Fundamental e da 3ª série do Ensino Médio. Nas duas primeiras edições (1990 e 1993) foram avaliados os alunos da 1ª, 3ª, 5ª e 7ª séries do Ensino Fundamental. As amostras são representativas de todos os estados da federação (apenas em 1990 participaram 25 estados e, em 1993, 26 estados).

Além de medir o desempenho escolar, o SAEB coleta dados sobre os alunos (por meio de questões socioeconômicas, culturais e da prática escolar), sobre os diretores (perfil e prática da gestão), os professores (perfil e práticas pedagógicas) e sobre a infra-estrutura dos estabelecimentos de ensino.

A análise dos resultados dos levantamentos do SAEB permite acompanhar a evolução do desempenho dos alunos e dos diversos fatores incidentes na qualidade e na efetividade do ensino ministrado nas escolas, possibilitando a definição de ações voltadas para a correção das distorções identificadas e o aperfeiçoamento das práticas e dos resultados apresentados pelas escolas.

Franco (2001) argumenta que é difícil oferecer explicações sobre os fatores que influenciam o aprendizado dos alunos por meio dos dados do SAEB porque o exame testa apenas uma vez cada aluno. Segundo o autor, a medida de proficiência é resultante do aprendizado dos alunos ao longo de muitos anos e não pode ser explicada em função do passado recente. Haveria necessidade de dupla medida de

proficiência para que, com base numa medida de proficiência prévia, fosse possível filtrar o aprendizado do aluno. O autor identifica este problema como sendo a principal limitação do SAEB.

Por ter caráter censitário, a Prova Brasil permite obter resultados médios de desempenho para Brasil, regiões, unidades da federação, municípios e escolas participantes. Logo, expande o alcance dos resultados do SAEB que, por ter caráter amostral, não podem ser calculados por município e por escola participante.

Ao enfocar o uso dos resultados do SAEB, Cotta (2001) indica a dificuldade de analisar o desempenho dos alunos de determinado Estado, uma vez que não há como separar a influência das políticas estaduais das iniciativas dos gestores municipais e das ações no âmbito da escola. É evidente que o gestor estadual não tem governabilidade sobre todas as ações que incidem sobre a qualidade do ensino ministrado. De acordo com a autora, esta situação paradoxal faz com que os usuários das informações de avaliação não se reconheçam nos resultados obtidos e impede a responsabilização dos gestores públicos perante a sociedade. A Prova Brasil já oferece a possibilidade de fazer análises para a esfera municipal e até mesmo de medir o desempenho por escola, o que de certa forma resolve o impasse identificado pela autora. Fica claro que o SAEB é uma avaliação diagnóstica e como tal não promove a *accountability*. A responsabilização de professores, gestores e até dos próprios alunos pelo seu desempenho só pode ocorrer com resultados de avaliações universais que sejam amplamente divulgados e que possam ser calculados por escola. Conclui-se que a Prova Brasil é uma avaliação que gera *accountability*.

Além da Prova Brasil deve-se destacar a criação do Índice de Desenvolvimento da Educação Básica (IDEB) em 2005. O índice foi criado com o propósito de tornar possível o monitoramento de escolas com alunos de baixo desempenho.

Um sistema educacional que reprova sistematicamente seus estudantes fazendo com que grande parte deles abandone a escola antes do tempo não é desejável. Tampouco pode ser considerado satisfatório um sistema que aprove seus alunos ao mesmo tempo em que permite que estes saiam da escola sem ter adquirido o nível de conhecimento apropriado (FERNANDES, 2007).

Logo, fazia-se necessário ter um indicador que pudesse combinar as informações referentes ao desempenho dos alunos com as de fluxo escolar (promoção, repetência e evasão). O cálculo do índice se dá por meio da combinação das notas padronizadas da Prova Brasil (indicador de proficiência) e da taxa média de aprovação dos alunos (indicador de fluxo escolar). A Prova Brasil permite que se tenham dados não apenas para o Brasil e unidades da Federação, mas também para cada município e escola participante. Desta forma, o IDEB pode ser calculado por unidade da Federação, município e até mesmo por escola¹.

O resultado do IDEB é utilizado como critério para que as escolas sejam consideradas prioritárias para receber assistência técnica e financeira por meio do Plano de Desenvolvimento da Escola (PDE-Escola). O PDE – Escola é uma ferramenta gerencial que auxilia a escola a realizar melhor o seu trabalho: focalizar sua energia, assegurar que sua equipe trabalhe para atingir os mesmos objetivos e avaliar e adequar sua direção em resposta a um ambiente em constante mudança (MEC, 2005).

¹ Escolas em área rural não possuem IDEB por não terem sido avaliadas pela Prova Brasil. Há municípios cuja rede de ensino está em área rural e, por esta razão, também não possuem IDEB. Há algumas escolas em área urbana sem IDEB já que o exame somente foi aplicado em escolas com mais de 20 alunos nas turmas avaliadas.

Na seção anterior, apresentaram-se algumas críticas existentes na literatura com relação às avaliações educacionais. Destacou-se a necessidade de estabelecer uma ligação entre a avaliação e a ação, a falta de unidade entre os atores do sistema educacional e os modelos de divulgação dos resultados dos exames.

No caso brasileiro, os obstáculos a serem superados não são diferentes. Franco (2001) aborda a questão da divulgação dos resultados e ressalta a necessidade de fazer a “ponte”. O autor sugere como alternativa uma apresentação de exemplos pedagógicos.

Outra crítica com relação à divulgação dos resultados refere-se ao excesso de técnicas complexas na apresentação das informações. Os processos de modelagem e as estatísticas utilizadas são complexos, mas ao transmitir as análises para o público em geral se faz necessário utilizar uma linguagem menos técnica, mais acessível. Franco (2001) propõe a utilização de gráficos simples que possam ser interpretados sem conhecimentos específicos de métodos quantitativos para que se consiga atingir o público alvo.

Cotta (2001) argumenta que o principal motivo pelo qual a avaliação geralmente não corresponde às expectativas que gera é a presunção de que as condições para o aproveitamento deste tipo de informação estão dadas. Segundo a autora, ocorre justamente o contrário: a avaliação é uma atividade que, por suas próprias características, exige que essas condições sejam intencionalmente criadas, sob pena de comprometer todo o esforço despendido.

Horta Neto (2007) investigou o uso dos resultados do SAEB pelas escolas e secretarias do Distrito Federal. A partir dos relatos coletados, foi possível perceber que discutir o desempenho dos alunos tomando por base informações objetivas é um processo que ainda enfrenta resistências. Segundo o autor, mesmo aqueles que utilizavam os dados coletados, faziam-no como se a melhoria do desempenho dos alunos fosse um fim em si mesmo, sem atentar para os fatores escolares que pudessem estar por trás desse desempenho.

Fica evidente que, assim como demonstrado na seção anterior por meio da experiência de outros países, ainda existe o desafio de construir a ligação entre a avaliação e a sala de aula para que estes exames não fiquem apenas com a função de traçar diagnósticos e possam de fato contribuir para uma mudança no sistema educacional.

Esta seção apresentou as principais avaliações nacionais da educação básica. A próxima trará alguns comentários a respeito da realização de avaliações internacionais e fará um breve histórico da participação do Brasil nestes exames.

4. Avaliações Internacionais da Educação Básica

No mesmo período em que introduziram avaliações nacionais de larga escala, alguns países em desenvolvimento passaram a integrar o grupo de países que se submetem a avaliações internacionais. Estas avaliações mensuram o produto de diversos sistemas educacionais simultaneamente e tem como grandes vantagens, a possibilidade de cooperação técnica, numa área que exige constante aprimoramento metodológico e a possibilidade de fazer uma análise comparada de sistemas de diferentes países. A

participação em provas internacionais não é custosa e pode ser muito rentável se os resultados forem utilizados para reformar o currículo e desenvolver programas de formação docente (WOLFF, 2007).

A Associação Internacional para a avaliação do progresso educacional (IEA) tem realizado estudos comparados desde 1959. O primeiro estudo que aplicou provas com mesmo conteúdo simultaneamente em diversos países foi feito por esta organização e avaliou a qualidade da educação em doze países (ROSS & PAVIOT, 2006).

A primeira iniciativa importante nesta área da qual o Brasil fez parte, como país convidado, foi o Projeto WEI – *World Education Indicator*, promovido pela UNESCO/OCDE. O WEI não avalia alunos com exames padronizados. É um levantamento estatístico com indicadores que permitem traçar diagnósticos e comparações entre os países participantes.

Até então, a incursão do Brasil em comparações internacionais se limitava ao envio de informações parciais para o anuário de estatísticas educacionais da UNESCO. Segundo Castro (2000), o principal benefício auferido pelo País foi a apropriação de metodologia para a produção de indicadores com comparabilidade internacional.

O WEI teve uma edição temática em 2002 sobre o financiamento da educação em que 18 países² foram avaliados. O Brasil ficou em 15º lugar ao considerar o gasto por aluno na educação básica.

Nos anos de 2005, 2007 e 2008, o Brasil participou novamente do WEI. Em 2007 avaliou-se de modo geral o sistema educacional de 19 países com objetivo de medir o progresso alcançado após a avaliação anterior. A edição de 2008 foi um estudo temático que avaliou escolas primárias de onze países³ com objetivo de entender e monitorar os determinantes da qualidade e da equidade.

Desta forma, o exame coletou dados referentes ao background dos alunos, características demográficas, informações sobre os professores, o tempo de instrução, as técnicas de ensino utilizadas, características de dirigentes e infra-estrutura das escolas (UNESCO – UIS, 2008).

Ao considerar características de infraestrutura básica o Brasil se situou no grupo dos que mais tem escolas com acesso a rede de esgoto, água, energia elétrica e computadores, mas é o país que menos oferece acesso a primeiros socorros em caso de emergência. O relatório concluiu que o acesso à educação primária está praticamente universalizado em todos os países participantes. Mas, há alguns países que apresentam problemas de fluxo escolar. O Brasil é o país com maior percentual de repetentes do grupo e, juntamente com o Paraguai, conta com professores que possuem o menor grau de formação e com o menor número de horas aula por ano.

Após participar da primeira edição do WEI, o Brasil passou a fazer parte do anuário *Education at a Glance*, que apresenta os indicadores educacionais dos países-membros da OCDE e dos participantes do projeto WEI (a partir de 1998).

² Argentina, Brasil, Chile, China, Egito, Índia, Indonésia, Jamaica, Jordânia, Malásia, Paraguai, Peru, Filipinas, Rússia, Tailândia, Tunísia, Uruguai e Zimbábue.

³ Argentina, Brasil, Chile, Índia, Malásia, Paraguai, Peru, Filipinas, Sri Lanka, Tunísia e Uruguai.

Outras iniciativas para o desenvolvimento de sistemas integrados de indicadores educacionais foram patrocinadas pelos fóruns de cooperação bilateral dos quais o Brasil faz parte. A primeira delas surgiu em 1997, no âmbito do MERCOSUL, através da definição de um conjunto básico de indicadores educacionais, a partir dos quais os países da região pretendiam estabelecer um sistema comum de informações.

Além disso, a Cúpula das Américas definiu uma linha específica voltada para a Avaliação e Indicadores Educacionais no plano de ação aprovado em 1997. A incorporação desta linha de ação no Plano de Ação da Cúpula das Américas criou condições para a ampliação da experiência que vinha sendo desenvolvida pelo Laboratório Latino-Americano de Avaliação da Qualidade da Educação (LLECE) por iniciativa da Oficina Regional de Educação da UNESCO para a América Latina e o Caribe (OREALC). O primeiro estudo internacional comparado contou com a participação de 13 países da região⁴. Os testes do PERCE (Primeiro Estudo Regional Comparativo e Explicativo) tiveram como objetivo avaliar o desempenho dos alunos da 3ª e 4ª série do ensino fundamental nas disciplinas de Língua e Matemática.

O aprendizado dos alunos brasileiros foi medido através de provas elaboradas pela equipe da OREALC e adaptadas à realidade brasileira. Além das provas, foram também aplicados questionários aos alunos testados, aos pais, professores e aos diretores das escolas onde os alunos estudam, com o objetivo de identificar fatores associados ao desempenho escolar.

Os resultados da avaliação de linguagem do PERCE foram divididos de acordo com o nível de leitura e com a localização geográfica das escolas. Assim, os níveis de leitura foram classificados em leitura literal, leitura literal por meio de paráfrase e leitura inferencial; e a localização geográfica classificada em "megacidade", urbana e rural. Calculou-se um percentual mínimo dos alunos avaliados que deveria atingir um nível satisfatório para cada nível de leitura. Os alunos brasileiros, que estudam em escolas localizadas nas categorias "megacidade" e urbana, conseguiram atingir o percentual mínimo nos três níveis de leitura. No entanto, os alunos que estudam em escolas em área rural obtiveram resultados abaixo do mínimo esperado em todos os níveis de leitura.

Após o PERCE, o LLECE desenvolveu o Segundo Estudo Regional Comparativo Explicativo⁵ (SERCE). O exame avaliou o desempenho dos estudantes da 3ª e 6ª séries do Ensino Fundamental, nas disciplinas de Matemática, Linguagem e Ciências na América Latina e no Caribe. Além da avaliação de aprendizagens, o SERCE utilizou questionários para recolher informação sobre os estudantes e suas famílias, os docentes e as escolas, de forma a discernir quais fatores têm maior incidência na aprendizagem dos alunos. Segundo a UNESCO (2008), os dois exames realizados pelo LLECE são complementares, mas não é possível estabelecer comparação entre os resultados obtidos porque o SERCE contou com aprimoramentos, como um desenho matricial das provas, instrumento que permitiu avaliar um maior número de conteúdos abordados na região.

Além dos estudos citados deve ser ressaltada a participação do Brasil no PISA (Programa Internacional de Avaliação de Estudantes), coordenado pela OCDE. O PISA é uma avaliação internacional de habilidades e conhecimentos de jovens de 15 anos, que visa aferir até que ponto os alunos próximos do

⁴ Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, Honduras, México, Paraguai, Peru, República Dominicana e Venezuela.

⁵ Países participantes: Argentina, Brasil, Chile, Colômbia, Costa Rica, Cuba, Equador, El Salvador, Guatemala, México, Nicarágua, Panamá, Paraguai, Peru, República Dominicana e Uruguai, assim como o estado mexicano de Nuevo León.

término da educação obrigatória adquiriram conhecimentos e habilidades essenciais para a participação efetiva na sociedade (INEP, 2001). A primeira edição do exame ocorreu no ano de 2000 e teve a participação de 32 países⁶. O exame foi concebido como um programa periódico que fosse capaz de realizar, num ciclo de nove anos, três avaliações de desempenho. As disciplinas avaliadas seriam leitura, matemática e ciências respectivamente. Todas as avaliações compreendem exames das três disciplinas, mas com foco especial em uma delas. Assim, a edição do PISA de 2000 teve foco em leitura, a de 2003 em matemática e a de 2006 em ciências. Em 2009, um novo ciclo se iniciará e novamente o foco será em leitura.

No caso brasileiro, conforme apresentado neste estudo, os resultados ainda eram bastante limitados no que concerne à educação básica na década de 90. A matrícula inicial abrangia apenas 90% da coorte correspondente. Havia um represamento nos primeiros quatro ou cinco anos, com as repetências levando ao atraso e ao abandono antes de terminar as oito séries do ensino fundamental. Somente 30% dos alunos obtinham seu certificado de ensino fundamental.

Há alguns problemas ao juntar no mesmo exame alunos de distintos sistemas tendo como critério de seleção a idade. Nos países da OCDE, praticamente não há atraso escolar por esta razão optam por definir as amostras por faixas etárias. No caso do PISA, alunos em torno de 15 anos constituem-se a população-alvo. Nos países da OCDE, quase todos os alunos desta faixa etária estão cursando a mesma série, portanto, já no ensino médio.

No entanto, os países Sul-Americanos pertencem a um grupo de sistemas educacionais com problemas graves ao considerar o fluxo escolar (abandono, repetência etc). Essa diferença cria um problema de interpretação dos dados uma vez que há muitos alunos que não atingiram o nível de escolaridade onde grande parte do conteúdo estaria sendo ensinado. Portanto, espera-se que respondam perguntas cujas respostas ainda não foram sequer incluídas nos currículos já cursados.

Outro ponto que deve ser destacado é o modelo de cobrança do conteúdo. No PISA, os alunos recebem textos tipicamente que remetem a situações reais. Após a sua leitura, devem responder perguntas relativas ao que está escrito. Segundo o Relatório Nacional do PISA (INEP, 2001), a tendência dos alunos brasileiros é ler superficialmente o texto e responder pelo que acham que este deveria estar dizendo e não pelo que efetivamente está escrito.

Assim como no caso dos exames nacionais, existem vários estudos na literatura que avaliam o impacto dessas avaliações nos sistemas educacionais. A maior parte dos dados produzidos é usada na formulação de políticas, mas não afetam diretamente a sala de aula (KELLAGHAN, 2006).

Nzomo & Makuwa (2006) avaliam o impacto do SACMEQ⁷ nos sistemas educacionais da Namíbia e do Quênia. Nestes países, o SACMEQ serviu para identificar algumas falhas no sistema, como o alto número de faltas dos alunos no Quênia e as disparidades regionais no que concerne a provisão da educação na Namíbia. O estudo ressalta que no Quênia os resultados das avaliações nacionais foram muito diferentes dos resultados do SACMEQ em termos de proficiência. Isto fez com que o governo realizasse seminários

⁶ Alemanha, Austrália, Áustria, Bélgica, Brasil, Canadá, Coreia do Sul, Dinamarca, Espanha, Estados Unidos, Rússia, Finlândia, França, Grécia, Holanda, Hungria, Irlanda, Islândia, Itália, Japão, Letônia, Liechtenstein, Luxemburgo, México, Noruega, Nova Zelândia, Polônia, Portugal, Reino Unido, República Tcheca, Suécia e Suíça.

⁷ *Southern and Eastern Africa Consortium for Monitoring Education Quality*. É um dos principais exames de avaliação educacional comparada. O primeiro ciclo do exame foi de 1995 a 1999 e teve como foco habilidades de leitura. Houve novo ciclo de 2000 a 2004.

com educadores e gestores da educação e levou a revisão dos critérios de avaliação nacional. Neste caso houve impacto na sala de aula uma vez que os professores participaram do debate e se propuseram a preparar os alunos de acordo com os novos critérios de proficiência.

Linnakylä (2006) apresenta o impacto do PISA na Finlândia. Os resultados dos alunos finlandeses no PISA foram considerados excelentes e tido como referência para vários países do mundo. Assim como no Quênia, os resultados do exame internacional divergiram dos resultados da avaliação nacional. No entanto, na Finlândia o resultado internacional foi bem mais satisfatório do que o nacional. Segundo o autor, o governo finlandês divulgou os resultados das duas avaliações simultaneamente e deu mais ênfase a avaliação nacional. Foi feito um plano para melhorar os indicadores educacionais que não foram satisfatórios. Ao mesmo tempo em que o governo se preocupava com o desempenho insatisfatório, jornais de todo o mundo enalteciam o ensino na Finlândia. Assim como no Quênia houve uma revisão dos critérios de avaliação do sistema nacional com objetivo de ter resultados semelhantes. No caso brasileiro, há similaridade entre os resultados do ENEM e os do PISA no tocante a proficiência em leitura.

5. Comentários Finais

A Avaliação em Larga Escala é uma realidade em diversos países, mas ainda precisa ser aprimorada para que se obtenham resultados mais efetivos. O diagnóstico por si só não é suficiente para mudar uma determinada situação. Embora o Brasil tenha avançado muito na coleta de dados e nos sistemas de avaliação ainda é preciso construir mecanismos para que os resultados sejam utilizados por gestores e professores de modo a melhorar a qualidade do ensino oferecido.

Referências

- BRASIL. Lei de Diretrizes e Bases da Educação Nacional, Lei n. 9.394, de 24 de dezembro de 1996.
- BRASIL. MEC/INEP. Relatório Nacional do PISA 2000. INEP Brasília 2001
- CASTRO, M.H. A participação do Brasil em Estudos e Avaliações Comparados Internacionais. Disponível em: <http://www.mre.gov.br/dc/textos/revista7-mat21.pdf>
- COTTA, T. Avaliação Educacional e políticas públicas: a experiência do sistema nacional de avaliação da educação básica (SAEB). Revista do Serviço Público. Ano 52 nº4 Brasília 2001.
- FERRER, G. Sistemas de evaluación de aprendizajes en America Latina: Balances y Desafios. Ed. PREAL Santiago Chile 2006.
- FRANCO, C. O SAEB – Sistema de Avaliação da Educação Básica: potencialidades, problemas e desafios. In: Revista Brasileira de Educação. n.17 pg 127-133. 2001
- HORTA NETO, J. Limites e possibilidades de utilização dos resultados do SAEB na gestão do sistema público de ensino: o caso do Distrito Federal. In: I Colóquio Ibero Americano de Política e Administração Escolar, 2007, Porto Alegre.
- _____. Um olhar retrospectivo sobre a avaliação externa no Brasil: das primeiras medições em educação até o SAEB de 2005. In: Revista Iberoamericana de Educación. Organización de Estados Iberoamericanos para La educación, ciência y cultura (OEI), nº. 42/5 Espanha 2007a
- IAIES, G. Evaluar las evaluaciones. In: Evaluar las Evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa. IIPE – UNESCO Buenos Aires Argentina 2003.
- INEP – Instituto Nacional de Estudos e Pesquisas Educacionais. Relatório SAEB 2005. Brasília, DF INEP/MEC 2006
- _____. PISA 2006. Brasília, DF: INEP/ MEC Disponível em: <http://www.inep.gov.br/download/internacional/pisa/PISA2006-resultados_internacionais_resumo.pdf>.

- KELLAGHAN, T. & GREANEY, V. Monitoring the Learning outcomes of Education Systems. World Bank Documents. Washington EUA 1996.
- ___ . Assessing Student Learning in Africa. World Bank Documents. Washington EUA 2004.
- LINNAKYLÄ, P. How can a country manage the impact of "excellent" cross-national research results? A case study from Finland. In: Cross-national studies of the quality of education. Org: ROSS, K. & GENEVOIS, I. J. UNESCO / IPE Paris França 2006
- MATTHEWS, P. Twelve Outstanding Secondary Schools – Excelling against the odds. OFSTED Inglaterra 2009 Disponível em: <http://www.ofsted.gov.uk/Ofsted-home/Twelve-outstanding-secondary-schools-Excelling-against-the-odds>
- Ministério da Educação (MEC). Instituto Nacional de Estudos e pesquisas Educacionais (INEP). Saeb 2001 Relatório Nacional. Brasília: INEP, 2001.
- Ministério da Educação (MEC). Instituto Nacional de Estudos e pesquisas Educacionais (INEP). Saeb 2001 Relatório Nacional – Novas Perspectivas. Brasília: INEP, 2002.
- Ministério da Educação (MEC). Instituto Nacional de Estudos e pesquisas Educacionais (INEP). Relatório Nacional Saeb 2003. Brasília: INEP, 2003.
- NZOMO, J. & MAKUWA, D. How countries move from cross-national research results to dissemination, and then to policy reform? In: Cross-national studies of the quality of education. Org: ROSS, K. & GENEVOIS, I. J. UNESCO / IPE Paris França 2006
- SAUS, J. Encuentros y Desencuentros con los procesos de evaluación de la calidad educativa em América Latina. In: Evaluar las Evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa. IPE – UNESCO Buenos Aires Argentina 2003.
- SOUSA, S. & OLIVEIRA, R. Sistemas de avaliação educacional no Brasil: característica, tendências e uso dos resultados. Relatório final. Universidade de São Paulo, Faculdade de Educação. São Paulo 2007.
- TEDESCO, J. Evaluar las evaluaciones. In: Evaluar las Evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa. IPE – UNESCO Buenos Aires Argentina 2003.
- UEMA, L. SAEB: inclusão pela qualidade. 2003. Disponível em:
<http://www.tvebrasil.com.br/salto/boletins2003/saeb/index.htm>
- UNESCO. Laboratório Latinoamericano de Evaluación de la Calidad de la Educación. Estudio cualitativo de escuelas con resultados destacables en siete países latinoamericanos. Santiago de Chile, 2002.
- ___ Primer estudio internacional comparativo. Santiago de Chile, 1998.
- ___ . Primer estudio internacional comparativo, segundo informe. Santiago de Chile, 2000.
- WOLFF, L. Los costos de las evaluaciones de aprendizaje en America Latina. In: Documentos PREAL n 38. Santiago Chile 2007.