



Pesquisas do Banco Mundial sobre Avaliações de Desempenho Educacional

Vincent Greaney | Thomas Kellaghan



VOLUME 1

AVALIAÇÃO DOS NÍVEIS DE DESEMPENHO EDUCACIONAL



BANCO MUNDIAL



Preencha a **ficha de cadastro** no final deste livro
e receba gratuitamente informações
sobre os lançamentos e as promoções da Elsevier.

Consulte também nosso catálogo
completo, últimos lançamentos
e serviços exclusivos no site
www.elsevier.com.br

Do original: *Assessing National Achievement Levels in Education*
Copyright © 2009, by The International Bank for Reconstruction and Development/The World Bank
Publicado originalmente em Inglês pelo Banco Mundial. Em caso de discrepâncias, prevalecerá a versão original em inglês.

Esta obra foi elaborada pela equipe do Banco Internacional para Reconstrução e Desenvolvimento/Banco Mundial. As afirmações, interpretações e conclusões aqui contidas não expressam necessariamente a opinião dos diretores executivos da instituição ou dos governos que eles representam.

O Banco Mundial não garante a exatidão dos dados incluídos nesta publicação. As fronteiras, cores, denominações e outras informações mostradas em qualquer mapa neste estudo não implicam nenhuma opinião da parte do Banco Mundial em relação à situação jurídica de qualquer território, bem como o endosso ou a aceitação de suas fronteiras.

This work was originally published by The World Bank in English as *National Assessments of Educational Achievement, Volume 1: Assessing National Achievement Levels in Education* in 2007. This Brazilian Portuguese translation was arranged by Elsevier Brazil. Elsevier Brazil is responsible for the quality of translation. In case of any discrepancies, the original language will govern.

The findings, interpretations, and conclusions expressed herein are those of the author(s) and do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent.

The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgement on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Direitos e permissões

O material contido nesta publicação é protegido por direito autoral. A cópia e/ou transmissão sem permissão de uma parte ou de todo o conteúdo poderão ser consideradas violação da lei aplicável. O Banco Internacional para Reconstrução e Desenvolvimento/Banco Mundial estimula a disseminação desta obra e normalmente permitirá com rapidez a reprodução de trechos deste relatório.

Para obter permissão de fotocópia ou reimpressão de qualquer parte deste livro, envie um pedido contendo informações completas para Copyright Clearance Center Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; telefone: 978-750-8400; fax: 978-750-4470; Internet: www.copyright.com.

Todas as outras perguntas sobre direitos e licenças, inclusive direitos subsidiários, devem ser dirigidas ao Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA.
Fax: 202-522-2422; e-mail: pubrights@Worldbank.org.

© 2011, The International Bank for Reconstruction and Development/The World Bank

Todos os direitos reservados e protegidos pela Lei nº 9.610, de 19/02/1998.

Copidesque: Shirley Lima da Silva Braz

Revisão: Edna Cavalcanti e Roberta Borges

Editoração Eletrônica: Estúdio Castellani

Elsevier Editora Ltda.

Conhecimento sem Fronteiras

Rua Sete de Setembro, 111 – 16º andar

20050-006 – Centro – Rio de Janeiro – RJ – Brasil

Rua Quintana, 753 – 8º andar

04569-011 – Brooklin – São Paulo – SP – Brasil

Serviço de Atendimento ao Cliente

0800-0265340

sac@elsevier.com.br

ISBN 978-85-352-3960-7

Edição original: ISBN 978-0-8213-7258-6

CIP-Brasil. Catalogação-na-fonte
Sindicato Nacional dos Editores de Livros, RJ

G825a Greaney, Vincent
Avaliação dos níveis de desempenho educacional / Vincent Greaney e Thomas Kellaghan ; tradução Maria Lúcia de Oliveira. – Rio de Janeiro : Elsevier, 2011.
il. – (Pesquisas do Banco Mundial sobre avaliações de desempenho educacional ; v. 1)
Tradução de: *Assessing national achievement levels in education*
Apêndice
Inclui bibliografia
ISBN 978-85-352-3960-7
1. Testes e medidas educacionais. 2. Avaliação educacional. I. Kellaghan, Thomas. II. Título. III. Série.

10-4613.

CDD: 371.262
CDU: 37.091.26



AGRADECIMENTOS

Uma equipe conduzida por Vincent Greaney (consultor do Grupo de Educação da Human Development Network do Banco Mundial) e Thomas Kellaghan (Educational Research Center, St. Patrick's College, Dublin) preparou esta série.

Também contribuíram para a série Sylvia Acana (Uganda National Examinations Board), Prue Anderson (Australian Council for Educational Research), Fernando Cartwright (Canadian Council on Learning), Jean Dumais (Statistics Canada), Chris Freeman (Australian Council for Educational Research), Hew Gough (Statistics Canada), Sara Howie (University of Pretoria), George Morgan (Australian Council for Educational Research), T. Scott Murray (DataAngel Policy Research) e Gerry Shiel (Educational Research Center, St. Patrick's College, Dublin).

O trabalho foi realizado sob a direção geral de Ruth Kagia, diretora do Setor de Educação do Banco Mundial, e Robin Horn, gerente do Setor de Educação. Robert Prouty iniciou e supervisionou o projeto até agosto de 2007. Marguerite Clarke supervisionou as últimas etapas, até a revisão e publicação. Somos gratos às contribuições do painel de revisão: Al Beaton (Boston College), Irwin Kirsch (Educational Testing Service) e Benoit Millot (Banco Mundial).

Comentários adicionais por pares foram feitos por diversos funcionários do Banco Mundial, incluindo Carlos Rojas, Eduardo Velez, Elizabeth King, Harry Patrinos, Helen Abadzi, Jee-Peng Tan, Marguerite Clarke, Maureen Lewis, Raisa Venalainen, Regina Bendokat, Robert Prouty e Robin Horn.

Agradecemos especialmente a Aidan Mulkeen e Sarah Plouffe. Recebemos apoio valioso de Cynthia Guttman, Matseko Ramokoena, Aleksandra Sawicka, Pam Spagnoli, Beata Thorstensen, Myriam Waiser, Peter Winograd e Hans Wagemaker. Também somos gratos a Patricia Arregui, Harsha Aturupane, Luis Benveniste, Jean-Marc Bernard, Carly Cheevers, Zewdu Gebrekidan, Venita Kaul, Pedro Ravela e Kin Bing Wu.

Desejamos agradecer às seguintes instituições a permissão de reproduzir seu material: Examinations Council of Lesotho (Conselho de Exames do Lesoto), International Association for the Evaluation of Educational Achievement (Associação Internacional para Avaliação do Aproveitamento Escolar), National Center for Education Statistics of the U.S. Department of Education (Centro Nacional de Estatísticas de Educação do Departamento de Educação) (EUA), Organisation for Economics Co-operation and Development (Organização para a Cooperação e o Desenvolvimento Econômico) (OCDE) e Papua New Guinea Department of Education (Departamento de Educação de Papua Nova Guiné).

Hilary Walshe ajudou a preparar o manuscrito. O desenho, a edição e produção dos livros foram coordenados por Mary Fisk e Paola Scalabrin, do Escritório de Publicações do Banco Mundial.

O Irish Educational Trust Fund; o Bank Netherlands Partnership Program; o Educational Research Center, Dublin; e o Australian Council for Educational Research deram generoso apoio à preparação e publicação desta série.

A tradução desta série para o português só foi possível com o generoso apoio da Russia Education Aid for Development Trust Fund.



PREFÁCIO

Num discurso para marcar os primeiros 100 dias de sua presidência do Grupo Banco Mundial, Robert Zoellick delineou seis temas estratégicos para orientar o trabalho do Banco com vista à promoção de uma globalização inclusiva e sustentável. Um desses temas centrou-se no papel do Banco como “uma instituição singular e especial de conhecimento e aprendizagem... um *brain trust* de experiência aplicada”. Zoellick observou que, para cumprir esse papel, o banco deve “enfocar, contínua e rigorosamente, os resultados e a avaliação da efetividade”.

Esse desafio é maior na área de educação, em que um grande corpo de evidências empíricas associa educação e crescimento econômico e indica que o aumento das taxas de matrícula e de conclusão são condições necessárias, mas não suficientes, para a redução da pobreza. Em vez disso, a melhoria dos produtos do aprendizado – melhores níveis de conhecimento e de habilidades cognitivas dos alunos – é fundamental para aliviar a pobreza e elevar a competitividade econômica (e será crucial para sustentar os ganhos já obtidos no acesso à educação). Em outras palavras, o pleno potencial da educação em relação ao crescimento econômico só pode se concretizar se o ensino oferecido for de alta qualidade e se forem desenvolvidos os conhecimentos e as habilidades cognitivas dos alunos.

As evidências disponíveis indicam que é muito baixa a qualidade dos resultados do aprendizado nos países em desenvolvimento. Ao mesmo tempo, apenas alguns desses países monitoram sistematicamente esses resultados, seja realizando as próprias avaliações do aproveitamento dos alunos, seja participando de avaliações regionais ou internacionais. A falta desse tipo de informação sistemática sobre o aprendizado dos alunos no âmbito do sistema educacional como um todo torna difícil avaliar os níveis globais de aproveitamento e o desempenho relativo de determinados subgrupos e acompanhar a evolução do aproveitamento ao longo do tempo. Isso também torna mais difícil determinar a efetividade das políticas governamentais destinadas a melhorar os resultados nessas áreas e em outras.

Essa mudança do foco, saindo do acesso e passando a se concentrar no aproveitamento, é uma questão central para o Banco e os países-membros. Trata-se também de uma área que carece de ferramentas e recursos adequados às necessidades dos países em desenvolvimento. Esta série de livros editados por Vincent Greaney e Kellaghan Thomas contribui de maneira significativa para cobrir essa lacuna. A série foi concebida para tratar de muitos dos problemas envolvidos e fazer com que os resultados do aprendizado ganhem maior importância na agenda educacional de países de baixa renda. Ela ajudará os países a desenvolver sua capacidade de medir os níveis nacionais de aprendizado dos alunos de formas mais válidas, sustentáveis e sistemáticas. Espera-se que essa capacidade se traduza na elaboração de políticas baseadas em evidências e que levem à melhoria perceptível da qualidade do aprendizado. É um componente importante do esforço para alcançar a promessa real de que a educação esteja associada a economias dinâmicas.

Marguerite Clarke
Especialista Sênior em Educação



SUMÁRIO

| | | |
|----------|---|-----------|
| 1 | INTRODUÇÃO | 1 |
| 2 | AVALIAÇÕES NACIONAIS DE APROVEITAMENTO DE ALUNOS | 7 |
| | Quais são os principais elementos de uma avaliação nacional? | 12 |
| | Quais as diferenças entre avaliação nacional e exames públicos? | 14 |
| 3 | POR QUE REALIZAR UMA AVALIAÇÃO NACIONAL? | 17 |
| 4 | DECISÕES ENVOLVIDAS NUMA AVALIAÇÃO NACIONAL | 23 |
| | Quem deve dar orientações de políticas para a avaliação nacional? | 23 |
| | Quem deve realizar a avaliação nacional? | 25 |
| | Quem aplicará as provas e os questionários? | 30 |
| | Qual o público-alvo a ser avaliado? | 31 |
| | A avaliação cobrirá todo o público-alvo ou apenas uma amostra? | 33 |
| | O que será avaliado? | 35 |
| | Como será avaliado o aproveitamento? | 40 |
| | Com que frequência serão realizadas as avaliações? | 45 |
| | Como deve ser divulgado o aproveitamento dos alunos? | 45 |
| | Que tipos de análises estatísticas devem ser feitas? | 48 |
| | Como devem ser apresentados e usados os resultados de uma avaliação nacional? | 50 |

| | | |
|------------------|---|------------|
| | Quais são os componentes de custo de uma avaliação nacional? | 52 |
| | Resumo das decisões | 55 |
| 5 | QUESTÕES ENVOLVIDAS NAS FASES DE PROJETO, IMPLEMENTAÇÃO, ANÁLISE, APRESENTAÇÃO DE RELATÓRIO E UTILIZAÇÃO DE UMA AVALIAÇÃO NACIONAL | 57 |
| | Projeto | 57 |
| | Implementação | 59 |
| | Análise | 61 |
| | Elaboração do relatório | 64 |
| | Divulgação e utilização das constatações | 65 |
| 6 | AVALIAÇÕES INTERNACIONAIS DO APROVEITAMENTO DE ALUNOS | 67 |
| | Crescimento da atividade de avaliação internacional | 69 |
| | Vantagens das avaliações internacionais | 72 |
| | Problemas das avaliações internacionais | 77 |
| 7 | CONCLUSÃO | 83 |
| APÊNDICES | | |
| A | ESTUDOS DE CASO DE PAÍSES | 91 |
| | A.1. Índia | 91 |
| | A.2. Vietnã | 93 |
| | A.3. Uruguai | 97 |
| | A.4. África do Sul | 99 |
| | A.5. Sri Lanka | 102 |
| | A.6. Nepal | 105 |
| | A.7. Chile | 107 |
| | A.8. Estados Unidos | 110 |
| | A.9. Uganda | 113 |
| B | ESTUDOS INTERNACIONAIS | 119 |
| | B.1. Tendências internacionais no estudo de matemática e ciências (TIMSS) | 119 |
| | B.2. Estudo internacional sobre o progresso do letramento em leitura (PIRLS) | 125 |
| | B.3. Programa internacional de avaliação de alunos – PISA | 129 |

| | | |
|----------|--|------------|
| C | ESTUDOS REGIONAIS | 139 |
| | C.1. Consórcio da África Austral e Oriental para monitoramento da qualidade da educação (SACMEQ) | 139 |
| | C.2. Programa de análise dos sistemas de educação da Confemen – PASEC | 148 |
| | C.3. Laboratório latino-americano de avaliação da qualidade da educação | 153 |
| | REFERÊNCIAS | 159 |
| | ÍNDICE | 169 |

QUADROS

| | | |
|-----|---|----|
| 2.1 | Etiópia: Objetivos da Avaliação Nacional | 11 |
| 2.2 | Exemplo de Questões Abordadas na Avaliação Nacional do Vietnã | 11 |
| 2.3 | Principais Elementos de uma Avaliação Nacional | 13 |
| 4.1 | Composição da CCN Proposta em Serra Leoa | 24 |
| 4.2 | Exemplos de Itens de Múltipla Escolha | 43 |
| 4.3 | Exemplos de Itens Abertos | 43 |
| 6.1 | A Experiência da África do Sul com Avaliações Internacionais | 82 |

FIGURAS

| | | |
|-----|---|----|
| 3.1 | O Hiato no Aproveitamento entre Alunos de 9 Anos nos Estados Unidos: Avaliação da Leitura, NAEP 1971-99 | 19 |
| 3.2 | Porcentagens de Alunos da 4ª Série nos Níveis “Proficiente” ou Acima em Leitura, NAEP 1992-2003 | 20 |
| 4.1 | Percentual Médio de Respostas Corretas em Matemática, por Áreas de Conteúdo no Lesoto | 47 |

| | | |
|-------|---|-----|
| A.9.1 | Distribuição das Pontuações do Teste de Letramento na 6ª Série em Uganda | 116 |
| B.3.1 | Amostra de Itens do Teste de Matemática do PISA | 132 |
| B.3.2 | Pontuações Médias de Letramento em Leitura e Pontuações na Subescala de Leitura do PISA, 2000 | 134 |
| B.3.3 | Níveis de Proficiência dos Alunos em Matemática no PISA | 135 |
| B.3.4 | Percentual de Alunos em Cada Nível de Proficiência na Escala de Matemática do PISA | 136 |
| B.3.5 | Percentual de Alunos em Cada Nível de Proficiência na Escala de Leitura do PISA | 137 |
| C.1.1 | Percentual de Alunos da 6ª Série que Alcançaram Níveis de Proficiência em Leitura no SACMEQ, 1995-98 | 146 |
| C.1.2 | Mudanças nas Pontuações de Letramento entre o SACMEQ I e o SACMEQ II | 147 |
| C.2.1 | Percentual de Alunos da 5ª Série com Baixo Aproveitamento, PASEC, 1996-2001 | 151 |
| C.3.1 | Percentual de Alunos em Cada Nível de Desempenho em Linguagem, por Tipo de Escola e Localização, LLECE 1997 | 157 |

TABELAS

| | | |
|-------|---|----|
| 2.1 | Diferenças entre Avaliações Nacionais e Exames Públicos | 16 |
| 4.1 | Opções para Realizar uma Avaliação Nacional | 26 |
| 4.2 | Vantagens e Desvantagens da Avaliação Baseada no Censo para Manter a Responsabilidade das Escolas | 35 |
| 4.3 | Processos de Compreensão de Leitura do PIRLS | 37 |
| 4.4 | Percentual que Alcançou a Meta ou o Nível de Domínio por Série, Connecticut, 2006 | 49 |
| 4.5 | Órgãos com a Responsabilidade Principal pelas Decisões numa Avaliação Nacional | 55 |
| 6.1 | Comparação entre o TIMSS e o PISA | 70 |
| 6.2 | Porcentagem de Alunos da 8ª Série que Alcançaram os Padrões de Referência Internacionais do TIMSS em Matemática: Países com Pontuações Altas e Baixas | 80 |
| A.2.1 | Porcentagens e Erros-Padrão de Alunos com Diferentes Níveis de Habilidade em Leitura | 96 |
| A.2.2 | Relação entre Algumas Variáveis Relativas ao Professor e ao Aproveitamento em Matemática | 96 |

| | | |
|-------|--|-----|
| A.5.1 | Dados sobre Antecedentes e Fonte na Avaliação Nacional do Sri Lanka | 103 |
| A.5.2 | Porcentagem de Alunos que Alcançaram o Domínio da Primeira Língua, por Província | 109 |
| A.7.1 | Índice para Prêmios de Merecimento para Escolas no Chile, 1998-99 | 110 |
| A.9.1 | Porcentagem de Alunos da 3ª Série em Uganda com Nível de Proficiência em Letramento em Inglês, 2005 | 114 |
| B.1.1 | Percentuais-alvos dos Testes de Matemática do TIMSS 2007 relativos aos Domínios de Conteúdo e Cognitivos, 4ª e 8ª Séries | 122 |
| B.1.2 | Distribuição do Aproveitamento em Matemática na 8ª Série no TIMSS | 124 |
| B.2.1 | Percentuais de Alunos da 4ª Série que Alcançaram os Padrões de Referência do PIRLS em Aproveitamento em Leitura | 128 |
| C.3.1 | Percentual de Alunos em Cada Nível de Desempenho em Linguagem, por Tipo de Escola e Localização, LLECE 1997 | 155 |
| C.3.2 | Percentual de Alunos em Cada Nível de Desempenho em Matemática, por Tipo de Escola e Localização, LLECE 1997 | 156 |



ABREVIACOES

| | |
|----------|---|
| CONFEMEN | Conferencia dos Ministros da Educao dos Pases de Lngua Francesa |
| DiNIECE | Direo Nacional de Informao e Avaliao da Qualidade da Educao – Argentina |
| IEA | Associao Internacional para Avaliao do Aproveitamento Escolar |
| IIEP | Instituto Internacional para o Planejamento Educacional |
| LLECE | Laboratrio Latino-Americano de Avaliao da Qualidade da Educao |
| MESyFOD | Programa de Modernizao da Educao Secundria e Formao de Docentes – Uruguai |
| NAEP | Avaliao Nacional do Progresso Educacional – Estados Unidos |
| NAPE | Avaliao Nacional do Progresso em Educao – Uganda |
| OCDE | Organizao para a Cooperao e o Desenvolvimento Econmico |
| PASEC | Programa de Anlise dos Sistemas Educacionais da CONFEMEN |

| | |
|--------|--|
| PIRLS | Estudo Internacional sobre o Progresso do Letramento em Leitura |
| PISA | Programa Internacional de Avaliação de Alunos |
| SACMEQ | Consórcio da África Austral e Oriental para Monitoramento da Qualidade da Educação |
| SIMCE | Sistema de Medição da Qualidade da Educação – Chile |
| SNED | Sistema Nacional de Avaliação do Desempenho dos Estabelecimentos Educacionais Subvencionados – Chile |
| SSA | Sarva Shiksha Abhiyan – India |
| TIMSS | Estudo das Tendências Internacionais no Estudo de Matemática e Ciências |
| UMRE | Unidade de Medição de Resultados Educacionais – Uruguai |
| UNEB | Comissão Nacional de Exames de Uganda |
| UNESCO | Organização das Nações Unidas para a Educação, a Ciência e a Cultura |

CAPÍTULO

1

INTRODUÇÃO

Neste livro introdutório, descrevemos as principais características das avaliações nacionais e internacionais que, nas décadas de 1990 e 2000, tornaram-se ferramentas extremamente populares para determinar a qualidade da educação. Esse aumento de popularidade reflete duas mudanças importantes. Em primeiro lugar, a crescente globalização e o interesse em mandatos globais, incluindo o Educação Para Todos (UNESCO, 2000). Em segundo, representa uma mudança geral de ênfase: o foco do processo de avaliação da qualidade da educação desvincula-se dos insumos (como taxas de participação dos alunos, instalações físicas, material didático e formação de professores) e passa a se concentrar nos resultados (como os conhecimentos e as habilidades que os alunos adquiriram em consequência da exposição ao ensino) (Kellaghan e Greaney, 2001b). Por sua vez, essa ênfase nos resultados pode ser considerada expressão da preocupação com o desenvolvimento do capital humano, a partir de duas convicções: de que o conhecimento está substituindo as matérias-primas e o trabalho como recurso para o desenvolvimento econômico e que a disponibilidade de conhecimentos e habilidades é essencial para determinar a taxa de desenvolvimento econômico de um país e sua competitividade num mercado internacio-

nal (Kellaghan e Greaney, 2001a). Para dar uma resposta a essa preocupação, foi necessário obter informações sobre o aproveitamento dos sistemas educacionais, e isso, por sua vez, resultou em que as provas tradicionais de aproveitamento, até então usadas para avaliar os alunos individualmente, passassem a ser utilizadas para se obter informação sobre os resultados do sistema educacional como um todo (ou de um segmento claramente definido).

O desenvolvimento da capacidade nacional de avaliação tem permitido que os ministérios de Educação – como parte de sua função gestora – descrevam os níveis nacionais de aproveitamento, especialmente nas principais áreas temáticas, e comparem os níveis de aproveitamento dos principais subgrupos (tais como meninos e meninas, grupos étnicos, alunos de escolas rurais e urbanas, alunos de escolas públicas e particulares). Adicionalmente, isso fornece evidências que permitem aos ministérios apoiarem ou refutarem afirmações de que, ao longo do tempo, os padrões de aproveitamento dos alunos estão subindo ou baixando.

Apesar do aumento das atividades de avaliação nos âmbitos nacional e internacional, há ainda, em vários lugares, falta de compreensão a respeito do valor potencial dos dados que a avaliação pode fornecer, bem como carência daquelas habilidades necessárias para realizar uma sólida avaliação técnica. Mesmo quando os países procedem a uma avaliação nacional ou participam de uma avaliação internacional, com frequência as informações resultantes não são plenamente exploradas. Existem diversas possíveis razões para isso: o envolvimento dos formuladores de políticas com a avaliação pode ter sido apenas periférico, e talvez não tenha se comprometido totalmente com ela; os resultados da análise podem não ter sido comunicados de forma inteligível para os formuladores de políticas; ou os formuladores talvez não tenham percebido plenamente as implicações das conclusões para a política social em geral ou para a política educacional, em especial no que se refere a definições sobre currículos, alocação de recursos, práticas de ensino e desenvolvimento profissional dos professores.

Para abordar essas questões, esta série de livros foi concebida como uma forma de levar ao conhecimento dos leitores a complexa tecnologia que se tem desenvolvido no campo da realização de avaliações nacionais

e internacionais. Este livro introdutório descreve os principais conceitos de avaliações e procedimentos nacionais e destina-se, principalmente, a formuladores de políticas e tomadores de decisões na área da educação. Os objetivos e as principais características das *avaliações nacionais* são descritos no Capítulo 2 (veja também Apêndice A). As razões para a realização de uma avaliação nacional são consideradas no Capítulo 3, e as principais decisões envolvidas no projeto e no planejamento de uma avaliação são abordadas no Capítulo 4. As questões que devem ser levadas em conta, bem como os erros comuns a serem evitados nas fases de projeto, implementação, análise, apresentação de relatórios e utilização de uma avaliação nacional são identificados no Capítulo 5. No Capítulo 6 estão descritas as *avaliações internacionais* de aproveitamento dos alunos, cujos procedimentos têm muitas características em comum com as avaliações nacionais (tais como amostragem, aplicação das provas, coleta de dados sobre antecedentes e contextos, e métodos de análise (veja Apêndice B).

A principal diferença entre as avaliações nacionais e internacionais destaca tanto o ponto forte quanto o ponto fraco de uma avaliação internacional. Ela tem como ponto forte o fato de fornecer dados sobre certo número de países, permitindo a cada um deles comparar os resultados de seus alunos com os resultados obtidos pelos de outros. O ponto fraco é que a exigência de que os instrumentos de teste sejam aceitáveis para todos os países participantes faz com que talvez não reflitam exatamente a gama de aproveitamentos dos alunos de diferentes países.

Outra característica das avaliações internacionais é que levam os países participantes a coletar dados de âmbito nacional para fazer sua análise interna. Com isso, os dados coletados para o estudo internacional podem ser usados para o que seria, de fato, uma avaliação nacional. No entanto, essa prática tem seus problemas, e os dados obtidos desse modo podem ser menos adequados para a elaboração de políticas do que se tivessem sido coletados para uma avaliação nacional específica.

Um procedimento intermediário, que se situa entre as avaliações nacionais de cada país e os estudos internacionais em grande escala que abrangem todo o mundo, é o *estudo regional*. Nesse, vários países de uma região, que provavelmente compartilham diversas características socioeconômicas e culturais, colaboram num estudo (veja Apêndice C).

Outra variante é a *avaliação subnacional*, limitada a uma região (estado ou província) dentro de um país. Avaliações subnacionais foram realizadas em vários países de grandes dimensões (como Argentina, Brasil e Estados Unidos) para satisfazer a necessidades de informações locais ou regionais. Esses exercícios são relativamente independentes e diferem das avaliações nacionais, na medida em que os instrumentos e procedimentos aplicados variam entre as regiões, o que impede comparações inter-regionais diretas do aproveitamento dos alunos.

No capítulo final deste volume, são apresentadas algumas conclusões gerais e feitas considerações sobre condições relativas ao desenvolvimento e à institucionalização da capacidade de avaliação nacional e à utilização ótima dos resultados da avaliação. No final do livro, são descritas as principais características das avaliações nacionais em nove países (Apêndice A), seguidas de descrições de três estudos internacionais (Apêndice B) e três estudos regionais (Apêndice C).

Os livros subsequentes desta série fornecem detalhes sobre a concepção e a implementação de uma avaliação nacional. São livros destinados a pessoas diretamente envolvidas na elaboração de provas e questionários e na coleta, análise ou descrição de dados de uma avaliação nacional. Foram concebidos como uma introdução aos aspectos técnicos fundamentais e às habilidades básicas necessárias para a realização dessas tarefas.

O Volume 2, *Desenvolvimento de testes e questionários para avaliação do desempenho educacional*, tem uma parte dedicada à elaboração de cada um dos três principais instrumentos: testes de aproveitamento, questionários e manuais de aplicação de provas. A Parte I aborda a elaboração de testes de aproveitamento e do papel que o contexto de um teste e o esquema ou tabela de especificações desempenham na elaboração. Descreve o processo de redação de itens (ou questões) e dá exemplos de vários tipos, incluindo itens de múltipla escolha, resposta curta e questões abertas. Também descreve o processo de revisão dos itens, ou painel, que é uma atividade essencial para garantir a validade do conteúdo da prova. Inclui diretrizes para realizar pré-testes, selecionar os itens para a prova final e produzir a versão final de uma prova. A parte termina com uma breve exposição sobre o treinamento de pontuadores

ou avaliadores e sobre a pontuação manual das questões. A segunda parte descreve as etapas da elaboração de questionários: concepção de um questionário, redação dos itens, pontuação e codificação de respostas e associação entre os dados derivados do questionário e as pontuações de aproveitamento dos alunos. A parte final descreve a concepção e o conteúdo de um manual de aplicação, a seleção dos aplicadores de teste e o papel que lhes cabe.

O Volume 3 também está dividido em três partes. A Parte I concentra-se em questões práticas a serem abordadas na execução de um programa de avaliação nacional em larga escala. Abrange planejamento, orçamentação, contratação de pessoal, obtenção de instalações e equipamentos, contatos com as escolas, seleção de aplicadores de provas, embalagem e expedição, e garantia da segurança da prova. Esta parte também cobre os aspectos logísticos de pontuação de provas, limpeza de dados e elaboração de relatórios. A Parte II inclui um manual com todos os passos necessários para que as equipes de avaliação desenhem uma amostra nacional adequada. A limpeza e o gerenciamento de dados são abordados na parte final.

A primeira metade do Volume 4 trata da geração de dados em nível de item, utilizando tanto a abordagem da teoria clássica das provas quanto a da teoria da resposta ao item. Os tópicos abordados incluem análise dos itens da prova-piloto e da prova final, monitoramento de mudanças no aproveitamento ao longo do tempo, construção de um teste a partir de itens previamente criados, calibragem e estabelecimento de níveis de aproveitamento ou de proficiência. A segunda metade do livro foi projetada para ajudar os analistas a realizar uma análise de nível básico dos resultados da avaliação nacional e inclui seções sobre medidas de tendência central e dispersão, diferenças na pontuação média, identificação de alunos de alto e baixo aproveitamento, correlação, regressão e representação visual de dados.

O uso dos resultados da avaliação do desempenho educacional, o Volume 5, centra-se nas formas de elaborar relatórios destinados aos que têm influência sobre as políticas. Apresenta uma metodologia para projetar uma estratégia de comunicação e divulgação de um programa nacional de avaliação. Também descreve a preparação de relatórios técnicos, de

comunicados de imprensa, de sessões de informação para os principais formuladores de políticas e de relatórios para professores e outros grupos de especialistas. A segunda parte do livro destaca alguns dos usos efetivos que os países têm dado aos resultados das avaliações nacionais, como elaboração de políticas, reforma curricular, distribuição de recursos, treinamento de professores, atribuição e cobrança de responsabilidades, monitoramento das mudanças registradas no aproveitamento e em outras variáveis ao longo do tempo.

Antecipa-se que as pessoas que estudarem o conteúdo desses livros e realizarem os exercícios propostos adquirirão as habilidades básicas necessárias para uma avaliação nacional. Devem, no entanto, ter em mente três fatores. Primeiro, não devem encarar os livros como um conjunto de meras fórmulas ou algoritmos a serem aplicados mecanicamente. Em vez disso, devem estar preparadas para tomar decisões em diversas etapas da avaliação nacional (por exemplo, na seleção do conteúdo da prova, na amostragem e na análise). À medida que se ganhe experiência, o processo de tomada de decisões sobre essas questões deverá ser simplificado. Em segundo lugar, os usuários podem, por vezes, solicitar o aconselhamento de profissionais mais experientes antes de tomar suas decisões. Em terceiro lugar, devem estar preparados para se adaptar às mudanças no conhecimento e na tecnologia que, inevitavelmente, ocorrerão nesta área nos próximos anos.

CAPÍTULO

2

AVALIAÇÕES NACIONAIS DE APROVEITAMENTO DE ALUNOS

Começamos o capítulo definindo uma avaliação nacional e listando questões a que se destinaria a responder. Em seguida, fazemos uma lista dos principais elementos de uma avaliação nacional. Finalmente, consideramos as diferenças entre avaliação nacional e exames públicos.

Uma avaliação nacional é projetada para descrever o aproveitamento dos alunos em uma área de currículo e agregar essas informações para fornecer uma estimativa do nível de aproveitamento no sistema educacional como um todo em determinada idade ou nível de ensino. Ela fornece dados para se realizar determinado tipo de auditoria nacional da educação, com vista a levar aos formuladores de políticas informações sobre aspectos-chave do sistema. Em geral, isso envolve a aplicação de provas de aproveitamento, a uma amostra ou a uma população de alunos, geralmente centrada num setor particular do sistema (como alunos da 5ª série ou alunos com 13 anos). Os professores e outros (por exemplo, pais, diretores e alunos) podem ser solicitados a fornecer informações sobre antecedentes e contextos, geralmente em questionários; quando relacionadas ao aproveitamento do aluno, essas informações podem permitir a percepção de como o aproveitamento está relacionado

com fatores como características das famílias, níveis de formação dos professores, atitudes dos professores com relação a diferentes áreas de currículos, conhecimento dos professores e disponibilidade de material de ensino-aprendizagem.

Os sistemas de avaliação nacional em várias partes do mundo tendem a apresentar traços comuns. Todos incluem avaliação da linguagem dos alunos, ou letramento em leitura, e de suas habilidades matemáticas (ou letramento em matemática). Alguns sistemas avaliam o aproveitamento dos alunos em um segundo idioma, em ciências, arte, música ou estudos sociais. Praticamente todos os sistemas nacionais de avaliação avaliam os alunos do primeiro grau. Em muitos sistemas, avaliações nacionais também são realizadas no segundo grau, geralmente no período de escolaridade obrigatória.

Os sistemas nacionais de avaliação também diferem de um país para outro. Em primeiro lugar, diferem na frequência com que são realizadas as avaliações. Em alguns países, a avaliação é realizada todos os anos, embora a área de currículo avaliada possa variar de ano para ano. Em outros sistemas, as avaliações são menos frequentes. Em segundo lugar, diferem quanto ao órgão que realiza a avaliação. Em alguns sistemas, isso cabe ao Ministério da Educação; em outros, cabe a um centro de pesquisa nacional, um consórcio de entidades educacionais, uma universidade ou uma comissão de exames. Em terceiro lugar, a participação de uma escola pode ser voluntária ou obrigatória. Quando voluntária, a não participação de algumas escolas quase invariavelmente distorce os resultados e leva a uma imagem inexata dos níveis de aproveitamento no sistema educacional.

Embora a maioria dos países industrializados já tivesse sistemas de avaliação nacional havia algum tempo, somente na década de 1990 a capacidade de gerir as avaliações tornou-se mais amplamente disponível em outras partes do mundo. Por exemplo, ocorreu um rápido desenvolvimento da criação de avaliações nacionais durante a década de 1990 em municípios da América Latina e do Caribe, muitas vezes para fornecer dados de linha base para reformas educacionais (Rojas e Esquivel, 1998). Depois da Declaração de Jomtien (veja *Declaração Mundial sobre Educação Para Todos*, 1990), esse desenvolvimento representou uma

mudança na avaliação da qualidade, tirando o foco dos insumos educacionais e passando a enfatizar resultados. O artigo 4º da Declaração de Jomtien afirma que o foco da educação básica deve estar “na aquisição do aprendizado e de seus resultados efetivos, e não mais exclusivamente na matrícula, na participação continuada em programas estabelecidos e no atendimento aos requisitos de certificação” (*Declaração Mundial sobre Educação Para Todos*, 1990, 5). Mais recentemente, o Plano de Ação de Dacar (UNESCO, 2000), produzido no final dos 10 anos de acompanhamento de Jomtien, também destacou a importância dos resultados do aprendizado. Entre a lista dos sete objetivos acordados, estava o de, até 2015, melhorar “todos os aspectos da qualidade da educação... de modo que resultados reconhecidos e mensuráveis sejam alcançados por todos, especialmente em letramento em linguagem, letramento em matemática e habilidades essenciais para a vida” (UNESCO, 2000, iv, 7).

Esses objetivos implicam que, para os países que se comprometeram a atingir as metas estabelecidas em Educação Para Todos, os esforços para melhorar a qualidade da educação terão de ser acompanhados de procedimentos que forneçam informações sobre o aprendizado dos alunos. Como resultado, os governos nacionais e os órgãos doadores têm aumentado bastante o apoio a avaliações nacionais destinadas a monitorar o aproveitamento dos alunos. O pressuposto frequente é que não só as avaliações nacionais fornecerão informações sobre o estado da educação, mas também que o uso dessa informação deve levar à melhoria do aproveitamento.

Se essa melhoria finalmente acontecerá ou não, é algo ainda a ser visto. Até agora, não parece haver se materializado a expectativa de que Educação Para Todos e o acompanhamento regular dos níveis de aproveitamento resultassem em melhoria dos padrões de aprendizado (Postlethwaite, 2004). Isso pode se dever ao fato de que, embora Educação Para Todos tenha levado ao rápido aumento no número de alunos que frequentam a escola, esse aumento não foi acompanhado de mais recursos (especialmente de professores treinados). Além disso, as informações obtidas nas avaliações têm sido, muitas vezes, de má qualidade, e, mesmo quando isso não ocorre, não vêm sendo sistematicamente levadas em conta para a tomada de decisões.

Todas as avaliações nacionais procuram respostas para uma ou mais das seguintes perguntas:

- Como está o aprendizado dos alunos no sistema educacional (com referência a expectativas gerais, objetivos do currículo, preparação para continuar o aprendizado ou preparação para a vida)?
- Existem evidências de pontos fortes e fracos específicos no conhecimento e nas habilidades dos alunos?
- Existem subgrupos na população-alvo que apresentam baixo aproveitamento? Existem disparidades, por exemplo, entre o aproveitamento de meninos e meninas, alunos de escolas urbanas e escolas rurais, alunos de idiomas ou grupos étnicos diferentes, ou alunos em diferentes regiões do país?
- Quais os fatores associados ao aproveitamento dos alunos? Em que medida o aproveitamento varia de acordo com as características do ambiente de aprendizagem (por exemplo, recursos da escola, formação e competência dos professores, tipo de escola) ou com as circunstâncias das famílias dos alunos e de suas comunidades?
- Estão sendo cumpridos os padrões estabelecidos pelo governo para a provisão de recursos (por exemplo, livros didáticos, qualificação dos professores e outros insumos relacionados à qualidade)?
- O aproveitamento dos alunos muda ao longo do tempo? Essa questão pode ser de particular interesse se estiverem sendo realizadas reformas no sistema educacional. Para responder à pergunta, é necessário realizar avaliações que produzam dados comparáveis em diferentes pontos no tempo (Kellaghan e Greaney, 2001b, 2004).

A maioria dessas questões foi levada em conta durante o processo de desenho e implementação da avaliação nacional feita pela Etiópia (veja o Quadro 2.1).

Uma característica da abordagem do Vietnã à avaliação nacional, além de avaliar o aproveitamento dos alunos, foi o forte foco em insumos fundamentais, como condições físicas das escolas, acesso a materiais educativos e qualificação dos professores (veja o Quadro 2.2).

QUADRO 2.1**Etiópia: Objetivos da Avaliação Nacional**

1. Determinar o nível de aproveitamento acadêmico dos alunos e o desenvolvimento de atitudes na educação básica etíope.
2. Analisar as variações no aproveitamento dos alunos por região, gênero, localização e idioma de instrução.
3. Explorar os fatores que influenciam o aproveitamento dos alunos no ensino básico.
4. Monitorar a melhoria do aproveitamento dos alunos desde a realização do primeiro estudo de base em 1999/2000.
5. Desenvolver a capacidade do sistema educacional para realizar avaliações nacionais.
6. Criar dados de linha de base confiáveis para o futuro.
7. Gerar recomendações para a formulação de políticas destinadas a melhorar a qualidade da educação.

Fonte: Etiópia, Organização Nacional de Exames, 2005.

QUADRO 2.2**Exemplo de Questões Abordadas na Avaliação Nacional do Vietnã****Questões Relativas a Insumos**

- Quais as características dos alunos da 5ª série?
- Quais as condições de ensino nas salas de aula da 5ª série e nas escolas do primeiro grau?
- Qual o estado geral dos edifícios escolares?

Questões Relativas a Padrões de Facilidades Educacionais

- Foram cumpridas as normas do Ministério em relação a:
 - Tamanho da turma?
 - Mobiliário da sala de aula?
 - Qualificações dos membros da equipe?

Questões Relativas à Equidade dos Insumos para as Escolas

- Houve equidade na distribuição dos recursos entre as províncias e entre escolas de uma mesma província, em termos de:
 - Insumos de recursos materiais?
 - Insumos de recursos humanos?

QUADRO 2.2**Questões Relativas ao Aproveitamento**

- Qual o percentual de alunos que atingiram os diferentes níveis de habilidade em leitura e matemática?
- Qual o nível atingido pelos professores da 5ª série em leitura e matemática?

Questões Relativas a Influências sobre o Aproveitamento

- Quais os principais fatores que explicam a variância no aproveitamento em leitura e matemática?
- Quais as principais variáveis em que se diferenciaram as escolas mais eficazes e as menos eficazes?

Fonte: Banco Mundial, 2004.

QUAIS SÃO OS PRINCIPAIS ELEMENTOS DE UMA AVALIAÇÃO NACIONAL?

Embora as avaliações nacionais possam variar na forma como são implementadas, tendem a conter vários elementos comuns (veja o Quadro 2.3 e Kellaghan e Greaney, 2001b, 2004).

É evidente, a partir da lista de elementos no Quadro 2.3, que é necessária uma boa dose de reflexão e preparação antes de se chegar à etapa em que os alunos estejam efetivamente realizando as provas de avaliação. Deve ser nomeado um órgão responsável pela coleta de dados, devem ser tomadas decisões sobre as questões de políticas a serem abordadas e as provas e os questionários devem ser elaborados e testados.

Na fase de preparação do teste propriamente dito, devem ser identificadas as amostras (ou o público-alvo) de escolas e alunos, as escolas devem ser contatadas e os aplicadores de prova, selecionados e treinados. Em alguns países (por exemplo, Índia, Vietnã e alguns países da África), os professores foram submetidos a uma avaliação que continha as mesmas tarefas dadas a seus alunos (veja A.1 e A.2 no Apêndice A e C.1 no Apêndice C). Após a aplicação da prova, será necessário dedicar muito tempo e esforço para preparar os dados para análise, avaliá-los e redigir os relatórios.

QUADRO 2.3**Principais Elementos de uma Avaliação Nacional**

- O Ministério da Educação nomeia um órgão implementador no âmbito do ministério ou um organismo externo e independente (por exemplo, um departamento universitário ou uma organização de pesquisa) e provê o financiamento.
- O Ministério da Educação determina necessidades de políticas que devem ser abordadas na avaliação, às vezes em consulta com as partes interessadas mais importantes na área da educação (por exemplo, representantes dos professores, especialistas em currículo, pessoas da área empresarial e pais de alunos).
- O Ministério da Educação, ou uma comissão de coordenação nomeada por ele, identifica o público-alvo a ser avaliado (por exemplo, alunos da 4ª série).
- O Ministério da Educação determina a área de aproveitamento a ser avaliada (por exemplo, letramento em linguagem ou em matemática).
- O órgão implementador define a área de aproveitamento e a descreve em função de conteúdos e habilidades cognitivas.
- O órgão implementador prepara provas de aproveitamento e material de apoio – questionários e manuais de aplicação – e adota medidas para garantir sua validade.
- As provas e os documentos de apoio são previamente testados pelo órgão implementador e, posteriormente, analisados pela comissão de coordenação e por outros órgãos competentes para determinar a adequação de currículos e assegurar que os itens levem em conta sensibilidades de gênero, étnicas e culturais.
- O órgão implementador seleciona a amostra-alvo (ou população-alvo) de escolas ou alunos, providencia a impressão de materiais e estabelece comunicação com as escolas selecionadas.
- O órgão implementador treina aplicadores de provas (por exemplo, professores, inspetores escolares ou estudantes universitários).
- Os instrumentos da pesquisa (provas e questionários) são aplicados nas escolas em uma data especificada e sob a direção geral do órgão implementador.
- O órgão implementador assume a responsabilidade pela coleta dos instrumentos da pesquisa e por encaminhá-los para a atribuição de pontos, além de se encarregar da limpeza e preparação dos dados para análise.
- O órgão implementador estabelece a confiabilidade dos instrumentos de avaliação e dos procedimentos.
- O órgão implementador procede à análise dos dados.
- Versões preliminares dos relatórios são preparadas pelo órgão implementador e revistas pela comissão de coordenação.

QUADRO 2.3

- Os relatórios finais são preparados pelo órgão implementador e divulgados pela autoridade competente.
- O Ministério da Educação e outros interessados relevantes revisam os resultados em função das necessidades de políticas a que pretendem atender e determinam um curso de ação adequado.

Fonte: Autores.

Ao tentar realizar uma avaliação nacional, os países de baixa renda têm de lidar com problemas de maior monta que os enfrentados por outros países. Seus orçamentos para educação podem ser exíguos. De acordo com dados de 2005 (Banco Mundial, 2007), alguns países alocam 2% ou menos do produto interno bruto para a educação pública (por exemplo, Bangladesh, Camarões, Chade, República Dominicana, Guiné, Cazaquistão, Laos, Mauritânia, Paquistão, Peru, República do Congo, Emirados Árabes Unidos e Zâmbia), em comparação com os mais de 5% da maioria dos países de renda média e alta.

Demandas concorrentes no setor de educação para atividades, como construção de escolas, formação de professores e fornecimento de materiais educativos, podem resultar na indisponibilidade de recursos para monitorar o aproveitamento escolar. Além disso, em muitos países de renda baixa, e até de renda média, a capacidade institucional para realizar uma avaliação nacional é limitada. Também podem ter de enfrentar outros problemas administrativos e de comunicação causados por estradas precárias ou serviços de correio e telefonia inadequados. Finalmente, por ser muito alta a variação nos níveis de aproveitamento de alunos de diferentes escolas em alguns países de baixa renda, é necessário que se trabalhe com uma amostra grande (veja UNEB 2006; Banco Mundial, 2004).

QUAIS AS DIFERENÇAS ENTRE AVALIAÇÃO NACIONAL E EXAMES PÚBLICOS?

Os exames públicos desempenham papel crucial em muitos sistemas educacionais quando se trata de certificar o aproveitamento dos alunos,

selecionar alunos que prosseguirão com os estudos e padronizar o que é ensinado e aprendido nas escolas. Às vezes, pensa-se que os exames públicos fornecem a mesma informação que uma avaliação nacional, o que pareceria eliminar a necessidade de um sistema nacional de avaliação num país que tenha um sistema de exame público. No entanto, tais exames não podem produzir o tipo de informação que uma avaliação nacional busca fornecer.

Em primeiro lugar, como os exames públicos desempenham papel importante na seleção de alunos (para o nível seguinte do sistema educacional e, às vezes, para emprego), buscam estabelecer diferença entre alunos com aproveitamento relativamente elevado e, por isso, talvez não façam uma cobertura adequada do currículo. Em segundo lugar, os exames, bem como as características dos alunos que se submetem a eles, mudam de ano para ano, limitando as inferências que poderiam ser feitas a partir de análises comparativas ao longo do tempo. Em terceiro lugar, o fato de fortes interesses estarem ligados ao desempenho (ou seja, de que o desempenho dos alunos num exame tenha consequências importantes para eles e talvez para seus professores) poderia levar os professores (e alunos) a se concentrarem nas áreas de currículo examinadas e a negligenciar outras áreas importantes que não são examinadas (por exemplo, habilidades práticas). Desse modo, o aproveitamento no exame não dá uma imagem exata do currículo pretendido. Embora existam algumas exceções, normalmente não se tomam decisões sobre alunos, professores ou escolas na sequência de um exame público.

Em quarto lugar, as informações sobre o aproveitamento dos alunos geralmente são exigidas numa idade anterior àquela em que são realizados os exames públicos. Em quinto lugar, o tipo de informação sobre antecedentes e contextos (o ensino, os recursos, os alunos e suas famílias) usada na interpretação dos dados de aproveitamento coletados em avaliações nacionais não está disponível para interpretar os resultados de exames públicos (Kellaghan, 2006). A Tabela 2.1 resume as principais diferenças entre avaliações nacionais e exames públicos.

TABELA 2.1

Diferenças entre Avaliações Nacionais e Exames Públicos

| | Avaliações nacionais | Exames públicos |
|--|--|--|
| Objetivo | Fornecer retornos para os formuladores de políticas. | Certificar e selecionar alunos. |
| Frequência | Para matérias específicas, oferecidas numa base regular (como a cada quatro anos). | Anualmente, e com mais frequência quando o sistema permite repetições. |
| Duração | Um ou dois dias. | Pode se estender por algumas semanas. |
| Quem é testado? | Normalmente, uma amostra de alunos de determinada série ou faixa etária. | Todos os alunos que desejem fazer o exame e estejam no nível de ensino correspondente. |
| Formato | Usualmente, múltipla escolha e respostas curtas. | Usualmente, redação e múltipla escolha. |
| O que está em jogo: importância para alunos, professores e outros | Pouca importância. | Grande importância. |
| Cobertura do currículo | Em geral, confinado a uma ou duas matérias. | Abrange as principais áreas temáticas. |
| Efeito sobre o ensino | Pouquíssimo efeito direto. | Principal efeito: professores tendem a ensinar o que se espera que caia no exame. |
| Reforço do ensino para alguns alunos | Muito improvável. | Frequentemente. |
| Os alunos recebem os resultados? | Raramente. | Sim. |
| São coletadas informações adicionais dos alunos? | Frequentemente, por meio de questionários respondidos por eles. | Raramente. |
| Pontuação | Em geral, envolve técnicas estatisticamente sofisticadas. | Normalmente, um processo simples, baseado num esquema predeterminado de marcação. |
| Efeito sobre o nível de aproveitamento do aluno | Improvável que tenha efeito. | Resultados fracos ou perspectiva de fracasso podem conduzir ao abandono precoce da escola. |
| Utilidade para acompanhar as tendências dos níveis de aproveitamento ao longo do tempo | Adequadas, se as provas forem concebidas já tendo em vista o acompanhamento. | Não são adequados, porque os itens dos exames e os públicos-alvo mudam de ano para ano. |

Fonte: Autores.

POR QUE REALIZAR UMA AVALIAÇÃO NACIONAL?

A decisão de realizar uma avaliação nacional pode decorrer de diversas razões. Frequentemente, as avaliações nacionais refletem os esforços de um governo para “modernizar” seu sistema educacional, introduzindo uma abordagem de gestão empresarial (corporativa) (Kellaghan, 2003). Essa abordagem baseia-se em conceitos utilizados no mundo dos negócios, como planejamento estratégico e foco em resultados, e pode envolver a responsabilização com base no aproveitamento. Vista dessa perspectiva, a avaliação nacional pode fornecer informações sobre um número limitado de medidas de resultado às quais atribuem importância os formuladores de políticas, os políticos e a comunidade educacional mais ampla.

O objetivo essencial desta abordagem é fornecer informações sobre o funcionamento do sistema educacional. Muitos governos não dispõem de informações básicas sobre aspectos do sistema – especialmente os níveis de aproveitamento dos alunos –, nem mesmo sobre insumos básicos. As avaliações nacionais podem proporcionar tais informações, o que representa uma condição prévia essencial para a elaboração de políticas sólidas. Por exemplo, a avaliação nacional do Vietnã ajudou a constatar que muitas salas de aula careciam de recursos básicos (Banco Mundial,

2004). De forma similar, a avaliação de Zanzibar revelou que 45% dos alunos não tinham um lugar para se sentar (Nassor e Mohammed, 1998). A avaliação nacional do Butão observou que alguns alunos tinham de passar várias horas a cada dia locomovendo-se entre a escola e sua casa (Butão, Câmara de Exames, Ministério da Educação, 2004). A avaliação da Namíbia revelou que muitos professores tinham pouco domínio de competências básicas em inglês e matemática (Makuwa, 2005).

Com o desenvolvimento da chamada “economia do conhecimento”, a necessidade de obter informações sobre o que os alunos aprendem na escola tem assumido importância crescente. Alguns analistas argumentam que os alunos necessitarão de níveis de conhecimentos e habilidades mais elevados que no passado, particularmente nas áreas de matemática e ciências, se quiserem participar, de forma significativa, do mundo do trabalho. Além disso, dado que o pronto acesso a bens e serviços aumenta com a globalização, considera-se que a capacidade de um país para competir com sucesso depende, em grande medida, das habilidades que tenham os trabalhadores e gestores para gerenciar a utilização de capital e tecnologia. Esse fator pode indicar a necessidade de se comparar o desempenho dos alunos em seu próprio sistema educacional com o desempenho de alunos em outros sistemas. No entanto, deve-se evitar o perigo de atribuir demasiada importância ao aproveitamento agregado dos alunos para explicar o crescimento econômico, já que existem muitos outros fatores envolvidos (Kellaghan e Greaney, 2001a).

As avaliações nacionais, quando realizadas por certo tempo, podem ser usadas para determinar se os padrões melhoraram, pioraram ou mantiveram-se estáveis. Muitos países em desenvolvimento enfrentam o problema de expandir as matrículas, construir muitas escolas e formar um grande número de professores enquanto tentam, ao mesmo tempo, melhorar a qualidade da educação – às vezes, no contexto de um orçamento diminuído. Nessa situação, os governos precisam monitorar os níveis de aproveitamento para determinar como as mudanças no número de matrículas e nas condições orçamentárias afetam a qualidade do aprendizado. Caso contrário, existe o risco de que o aumento nas taxas de matrícula possa ser facilmente aceito como prova de melhoria na qualidade da educação.

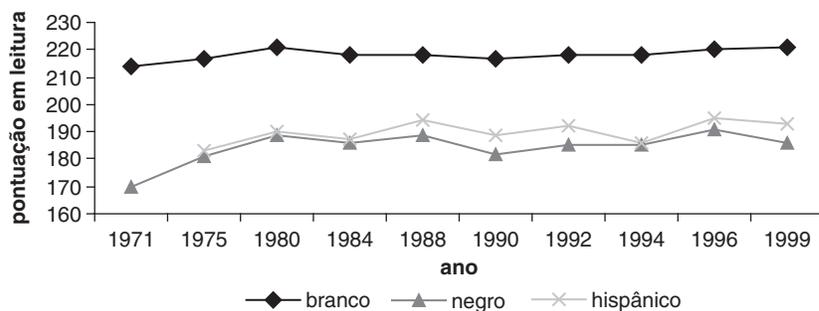
Os dados da avaliação nacional têm sido utilizados para monitorar o aproveitamento ao longo do tempo. Uma série de estudos feitos na África entre 1995/96 e 2000/01 revelou um declínio significativo nas pontuações de letramento em leitura no Malawi, na Namíbia e Zâmbia (veja a Figura C.1.2 no Apêndice C). Nos Estados Unidos, a Avaliação Nacional do Progresso Educacional (NAEP), que tem acompanhado os níveis de aproveitamento em leitura ao longo de quase três décadas, constatou que, embora as crianças negras e hispânicas de 9 anos tenham reduzido o hiato de aproveitamento que havia entre elas e as crianças brancas até cerca de 1980, o diferencial na pontuação das provas manteve-se relativamente constante nos anos seguintes (Figura 3.1). Também nos Estados Unidos, a NAEP ajudou a identificar as mudanças ocorridas nos níveis de aproveitamento em leitura em diversos estados (Figura 3.2).

No Nepal, os resultados das avaliações nacionais foram usados para monitorar mudanças no aproveitamento ao longo do período 1997-2001 e, em particular, os efeitos das decisões de políticas relativas a orçamento, currículos, livros didáticos, materiais didáticos e formação de professores (veja A.6 no Apêndice A).

Quando os dados de avaliações nacionais são usados para monitorar o aproveitamento ao longo do tempo, a mesma prova deve ser utilizada em cada avaliação ou, se forem utilizadas provas diferentes, alguns

FIGURA 3.1

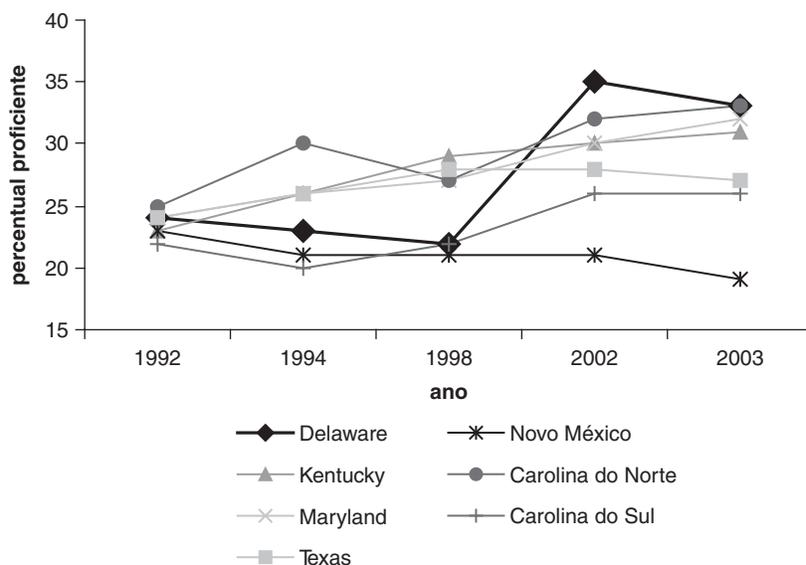
O Hiato no Aproveitamento entre Alunos de 9 Anos nos Estados Unidos: Avaliação da Leitura, NAEP 1971-99



Fonte: Winograd e Thorstensen, 2004.

FIGURA 3.2

Porcentagens de Alunos da 4ª Série nos Níveis “Proficiente” ou Acima em Leitura, NAEP 1992-2003



Fonte: Winograd e Thorstensen, 2004.

itens devem ser comuns, de modo que os desempenhos nas provas possam ser equiparados ou vinculados. Qualquer que seja o caso, os itens comuns devem ser mantidos em sigilo, para que o conhecimento do aluno ou do professor a respeito de seu conteúdo não invalide as comparações feitas.

Outros usos que podem ser feitos de uma avaliação nacional dependem de os dados terem sido coletados numa amostra de escolas ou num censo no qual se obtiveram informações sobre todas (ou quase todas) as escolas. Em ambos os casos, os resultados podem ser usados para fornecer orientação aos formuladores de políticas interessados em melhorar a qualidade da educação. Por exemplo, os resultados podem ajudar os governos a identificar a força da relação entre qualidade do aprendizado dos alunos e vários fatores sobre os quais têm algum controle (por exemplo, a disponibilidade de livros didáticos, o tamanho da turma e o número de anos anteriores de treinamento dos professores).

Uma análise dos resultados pode levar a decisões que afetem a destinação de recursos no sistema educacional em geral (por exemplo, para a reforma dos currículos e livros didáticos ou para formação de professores) ou em categorias de escolas com características específicas (por exemplo, escolas em áreas rurais ou escolas que atendem a alunos de áreas em condições socioeconômicas desfavoráveis). Podem ser encontrados muitos exemplos do uso dos resultados de avaliações nacionais e internacionais para tais fins. Têm sido usados na Austrália para criar programas destinados a aumentar a participação e o desempenho das meninas em matemática e ciências (Keeves, 1995); têm conduzido a reformas curriculares em países de baixa e média renda (Elley, 2005); ajudaram a reorientar recursos financeiros para as escolas mais pobres no Chile (veja A.7 no Apêndice A) e promoveram o profissionalismo dos professores no Uruguai (veja A.3 no Apêndice A).

Os resultados de uma avaliação nacional também podem ser usados para mudar a prática em sala de aula (Horn, Wolff e Vélez, 1992). No entanto, levar informações aos professores e promover alterações em seu comportamento com vista a aumentar substancialmente o aproveitamento dos alunos não é uma tarefa fácil. A pressão sobre as escolas e salas de aula para a mudança é maior quando os resultados de uma avaliação nacional se baseiam num censo escolar, e não numa amostra, e quando fortes interesses estão ligados ao desempenho. Nenhuma ação específica pode ser empreendida pelas autoridades além da publicação de informações sobre o aproveitamento (por exemplo, em tabelas classificativas). Mas podem-se associar sanções ao desempenho. As sanções podem assumir a forma de recompensa por melhor desempenho (por exemplo, as escolas, os professores ou ambos recebem incentivos econômicos se os alunos atingirem uma meta específica) ou de “punição” por um mau desempenho (por exemplo, não promoção dos alunos ou demissão de professores) (veja A.7 no Apêndice A para uma breve descrição do programa de recompensa do Chile).

Quando uma avaliação nacional obtém informações sobre o aproveitamento dos alunos em todas as escolas (ou na maior parte delas), alguns formuladores de políticas podem ver aí a oportunidade de usar esses dados para avaliar a qualidade dos professores e das escolas. Obviamente,

os professores e os alunos devem assumir alguma responsabilidade pelo aprendizado, mas o papel das instituições, dos órgãos e das pessoas que exercem controle sobre os recursos e as atividades das escolas também deve estar refletido em um sistema de responsabilização. Repartir equitativamente as responsabilidades de todas as partes interessadas é importante, seja no caso de uma avaliação baseada em amostra ou num censo. A avaliação nacional do Uruguai é um bom exemplo do reconhecimento da responsabilidade que cabe a cada um dos interessados (incluindo o Estado) pelo aproveitamento dos alunos (veja A.3 no Apêndice A).

Em alguns casos, uma avaliação nacional pode ter papel meramente simbólico destinado a legitimar a ação estatal: adotam-se modelos internacionalmente aceitos de modernidade e confere-se ao processo de decisão política uma aparência de racionalidade científica (Benveniste, 2000, 2002; Kellaghan, 2003). Quando é esse o fator que motiva uma avaliação nacional, o ato de avaliar tem maior importância que seus resultados. Se uma avaliação nacional é realizada apenas para atender à exigência de um órgão financiador, ou mesmo para cumprir os compromissos internacionais assumidos por um governo no sentido de monitorar os progressos destinados a alcançar os Objetivos de Desenvolvimento do Milênio, tal avaliação pode ter pouco mais que um valor simbólico, e suas conclusões podem não ter peso algum na gestão do sistema educacional ou na formulação de políticas.

CAPÍTULO

4

DECISÕES ENVOLVIDAS NUMA AVALIAÇÃO NACIONAL

Neste capítulo, consideramos 12 decisões envolvidas no planejamento de uma avaliação nacional (veja Greaney e Kellaghan, 1996; Kellaghan, 1997; e Kellaghan e Greaney, 2001b, 2004).

QUEM DEVE DAR ORIENTAÇÕES DE POLÍTICAS PARA A AVALIAÇÃO NACIONAL?

O Ministério da Educação deve nomear uma comissão de coordenação nacional (CCN) para fornecer orientação geral ao órgão que realizará a avaliação. A comissão pode ajudar a garantir que a avaliação nacional seja prestigiada e que sejam abordadas questões-chave de políticas que interessem ao ministério e a outros. Também poderia ajudar a resolver os graves problemas administrativos e financeiros que podem surgir durante a execução da avaliação nacional. Dar à CCN uma participação efetiva desde a fase de definição do rumo e do propósito da avaliação nacional também aumentará a probabilidade de que os resultados da avaliação venham a desempenhar papel relevante na formulação de políticas futuras.

A composição de uma CCN variará de país para país, dependendo da estrutura de poder dentro do sistema educacional. Além de representantes do Ministério da Educação, a CCN pode incluir representantes dos principais grupos étnicos, religiosos e linguísticos, bem como os grupos a cujos membros caberá tomar medidas em função dos resultados (como formadores de professores, professores, inspetores escolares e especialistas em currículo). O Quadro 4.1 apresenta uma lista de membros que deveriam integrar uma comissão de coordenação para uma avaliação nacional em Serra Leoa, proposta pelos participantes em um seminário internacional. O fato de se levarem em conta as necessidades de informação dos vários interessados deve contribuir para que o exercício de avaliação nacional não resulte em um relatório que será criticado ou ignorado devido à sua incapacidade de examinar as perguntas “corretas”.

QUADRO 4.1**Composição da CCN Proposta em Serra Leoa**

- Comissão de Educação Básica
- Movimento da Sociedade Civil
- Secretaria Descentralizada
- Diretor-Geral da Educação (presidente)
- Direção de Planejamento da Educação
- Conselho Inter-Religioso
- Centro Nacional de Desenvolvimento de Pesquisa de Currículos
- Sindicato dos Professores de Serra Leoa
- Estatísticas de Serra Leoa
- Faculdades de Formação de Professores
- Conselho de Exames da África Ocidental

A CCN não deve ser sobrecarregada com reuniões, nem deve ser chamada a lidar com as tarefas rotineiras de execução relacionadas com a avaliação nacional. Em alguns casos, pode fornecer orientação na fase inicial de identificação do objetivo e do fundamento lógico da avaliação,

determinando as áreas de currículos e os níveis de ensino que serão avaliados, ou selecionando o órgão (ou os órgãos) que conduzirá a avaliação – embora os itens de prova também possam ser decididos antes da criação da comissão. É provável que a CCN seja mais ativa no início do exercício de avaliação e que o órgão implementador se responsabilize pela maior parte do detalhamento do trabalho – como desenvolvimento de instrumentos, amostragem, análise e relatórios. O órgão implementador, no entanto, deve fornecer à CCN as versões preliminares das provas e dos questionários, bem como as descrições dos procedimentos propostos, de modo que os membros da comissão possam dar orientações e garantir que as necessidades de informação que deflagraram a avaliação estejam sendo atendidas adequadamente. Os membros da CCN também devem analisar as versões preliminares dos relatórios elaboradas pelo órgão implementador.

Responsabilidade de dar a orientação política: Ministério da Educação

QUEM DEVE REALIZAR A AVALIAÇÃO NACIONAL?

A avaliação nacional deve ser realizada por uma equipe ou organização que goze de credibilidade e cujo trabalho possa impor respeito e aumentar a probabilidade de ampla aceitação dos resultados. Vários países atribuíram a responsabilidade pelas avaliações nacionais a diferentes tipos de grupos, como, por exemplo, equipes criadas no âmbito do Ministério da Educação, organismos autônomos (universidades, centros de pesquisa) ou equipes técnicas estrangeiras. Seria de se esperar que diversos fatores influenciem essa decisão, incluindo os níveis de capacidade técnica nacional e as circunstâncias administrativas e políticas. A Tabela 4.1 lista algumas vantagens e desvantagens potenciais das diferentes categorias de organismos de implementação. Devem-se levar em conta esses itens no momento de decidir quem realizará a avaliação.

TABELA 4.1
Opções para Realizar uma Avaliação Nacional

| Órgão designado | Vantagens | Desvantagens |
|---|--|---|
| Extraído de pessoal do Ministério da Educação | Provavelmente conta com a confiança do ministério. Goza de fácil acesso ao pessoal-chave, a materiais e dados (por exemplo, dados sobre a população escolar). Talvez não seja necessário obter fundos para pagar esse pessoal. | Os resultados podem ser objeto de manipulação política, inclusive supressão. Pode ser visto com ceticismo por outras partes interessadas. Pessoal que pode estar incumbido de muitas outras tarefas. Pode faltar capacidade técnica. |
| Extraído de pessoal do órgão de exames públicos | Normalmente goza de confiança. Tem experiência com a execução de avaliações seguras. Talvez não seja necessário obter fundos para pagar esse pessoal. Algumas habilidades (por exemplo, o desenvolvimento de testes) que podem ser transferidas para reforçar a unidade responsável pelo teste. A probabilidade de ser sustentável é maior que em alguns outros modelos. | Pessoal que pode estar incumbido de muitas outras tarefas. A capacidade técnica pode ser fraca. Talvez não tenha acesso fácil aos dados. A experiência com exames públicos pode resultar em itens de teste muito difíceis. |
| Extraído do setor universitário/de pesquisa | Resultados podem ter maior credibilidade perante as partes interessadas. Maior probabilidade de haver alguma competência técnica. Talvez os dados sejam utilizados para estudos adicionais sobre o sistema de ensino. | É preciso levantar fundos para cobrir custos de pessoal. Pode ser menos sustentável que alguns outros modelos. Pode entrar em conflito com o Ministério da Educação. |

| | | |
|--|---|---|
| Recrutado como assistência técnica estrangeira (AT) | Maior probabilidade de que seja tecnicamente competente. Tipo de financiamento que pode ajudar a garantir a conclusão no tempo certo. | Probabilidade de ser caro. Pode não ser sensível ao contexto educacional. Dificuldade em garantir a sustentabilidade da avaliação. Possivelmente pouco reforço da capacidade nacional. |
| Composto por uma equipe nacional com algum apoio de AT internacional | Pode melhorar a capacidade técnica dos nacionais. Pode assegurar a realização no tempo certo. Pode aumentar a credibilidade dos resultados. | Possível dificuldade para coordenar os trabalhos dos membros da equipe nacional e da AT. Talvez seja difícil garantir a transferência de competências para os nacionais. |
| Equipe do ministério com apoio de AT nacional | Pode garantir apoio do ministério para obter AT nacional. Menos caro do que a AT internacional. | AT nacional que pode não ter a capacidade técnica necessária. Outras desvantagens potenciais listadas no item Ministério da Educação e que podem se aplicar ao caso. |

Fonte: Autores.

Em alguns casos, as tradições e a legislação podem impor restrições à liberdade que um Ministério da Educação tem para escolher um órgão implementador. Na Argentina, por exemplo, as províncias devem autorizar os conteúdos curriculares a serem aferidos na avaliação nacional. Inicialmente, pediu-se às províncias que produzissem os itens de teste; no entanto, muitas não tinham capacidade técnica para fazê-lo. Numa fase posterior, as províncias receberam um conjunto de possíveis perguntas para aprovação, e a Direção Nacional de Informação e Avaliação da Qualidade da Educação (DiNIECE) criou os instrumentos finais de avaliação a partir do conjunto de itens pré-aprovados. Mais recentemente, os itens das provas foram concebidos de modo independente pelo pessoal da universidade e aprovados pelo Conselho Federal. O DiNIECE permanece responsável pela concepção dos testes de aproveitamento, pelas análises dos resultados, bem como pela coordenação geral das atividades anuais de avaliação.

Vale a pena refletir sobre a grande variedade de habilidades necessárias para realizar uma avaliação nacional antes de decidir a quem deve ser atribuída a responsabilidade. Essa questão é abordada com mais detalhes no Volume 3 desta série. A avaliação nacional é, fundamentalmente, um esforço de equipe. A equipe deve ser flexível, estar disposta a trabalhar sob pressão e de forma colaborativa, e preparada para aprender novas abordagens de avaliação e novas tecnologias associadas. O líder da equipe deve ter grandes habilidades gerenciais. Ele terá a obrigação de organizar a equipe, coordenar e agendar atividades, apoiar o treinamento e obter e controlar os recursos financeiros. O líder da equipe deve ser politicamente astuto, porque terá de se reportar à CCN e atuar como elemento de ligação junto às autoridades nacionais, regionais e, em alguns casos, a órgãos de governo em nível local e a representantes das partes interessadas (tais como professores e entidades religiosas).

A equipe deve ter excelentes habilidades operacionais ou de execução. As tarefas a serem realizadas incluem oficinas para redatores de itens e aplicadores de provas; organização da impressão e distribuição de provas, questionários e manuais; contatos com escolas; desenvolvimento de material de treinamento, e coleta e gravação de dados. Será necessária uma pequena equipe especializada de elaboradores de testes para analisar o currículo, desenvolver quadros de especificações ou um

esquema de teste, fazer uma redação preliminar dos itens, selecionar os itens após o pré-teste ou piloto e aconselhar sobre a pontuação. Após a aplicação da prova, as perguntas abertas e de múltipla escolha deverão receber uma pontuação.

A equipe vai exigir o apoio de uma ou mais pessoas com competência estatística e analítica em vários tópicos: seleção de amostras, ponderação de dados, entrada de dados e preparação de arquivos, análise de itens dos dados das provas, análise estatística geral dos resultados globais e preparação de arquivos de dados para outros usuários (por exemplo, acadêmicos e alunos de pós-graduação) que realizarão análises secundárias. Muitos países em desenvolvimento carecem de capacidade nessa última área, levando a situações em que os dados são coletados, mas nunca devidamente analisados ou divulgados.

A equipe deve ter o pessoal necessário para redigir e divulgar os resultados, produzir comunicados de imprensa e panfletos ou boletins informativos direcionados a públicos específicos. Também seria de se esperar que a equipe desempenhasse papel fundamental na organização de oficinas para professores e outros funcionários da área de educação, para que possam discutir a importância dos resultados e suas implicações para o ensino e o aprendizado.

A maioria dos membros da equipe pode trabalhar em tempo parcial e ser empregada se e quando necessário. Essa categoria poderia incluir os redatores de itens – principalmente professores em atividade com bom conhecimento do currículo – e especialistas em amostragem e análise estatística. Os membros da equipe poderiam ser recrutados fora do setor de educação. Por exemplo, um organismo encarregado do censo nacional poderia ser uma boa fonte de competências na área de amostragem. Pessoal de computação com experiência relevante poderia ajudar na limpeza de dados, e jornalistas podem contribuir com a elaboração de comunicados de imprensa que despertem interesse. Nem o Camboja nem a Etiópia empregaram pessoal em tempo integral para realizar suas avaliações nacionais.

Responsabilidade de realizar as avaliações nacionais: Órgão implementador (Ministério da Educação, comissão de exames, órgão de pesquisa, universidades).

QUEM APLICARÁ AS PROVAS E OS QUESTIONÁRIOS?

As tradições administrativas nacionais e as percepções dos níveis de confiança, bem como as fontes de financiamento, tendem a influenciar a seleção do pessoal responsável pela aplicação de provas e questionários numa avaliação nacional. A prática varia. Por exemplo, alguns países têm usado alunos de pós-graduação, enquanto a Zâmbia envolveu inspetores escolares e funcionários graduados do ministério na aplicação da prova e do questionário. Outros países têm utilizado professores experientes que trabalham em outras escolas ou professores aposentados. Nas Maldivas, um aplicador de teste tem de fazer parte do quadro de pessoal de uma escola localizada numa ilha que não aquela em que está a escola-alvo.

Os aplicadores de teste devem ser cuidadosamente selecionados. Devem ter boas habilidades organizacionais, experiência de trabalho em escolas e estar empenhados em seguir rigorosamente as orientações da prova e do questionário. Idealmente, devem ter experiência em sala de aula, falar o mesmo idioma dos alunos, com o mesmo sotaque, e ter uma forma de ser que imponha autoridade, sem ser ameaçadora. O Volume 3 considera as vantagens e desvantagens de haver professores, inspetores, treinadores de professores, membros da comissão de exames e alunos universitários como aplicadores de testes.

Embora possa parecer administrativamente conveniente e muito econômico utilizar como aplicadores de provas os próprios professores dos alunos que estão participando da avaliação nacional, raramente isso é feito, por diversas razões. Alguns professores podem sentir que a eficácia de seu ensino está sendo avaliada. Outros podem ter dificuldade de abandonar a prática normal de tentar ajudar os alunos e talvez não consigam adaptar-se à abordagem formal de um teste. E outros podem fazer cópias das provas ou de partes dela, eliminando a possibilidade de se usarem esses itens em futuras avaliações nacionais. Usar os professores para aplicar as provas a seus próprios alunos também pode diminuir a percepção do público quanto à confiabilidade dos resultados da avaliação.

Responsabilidade de aplicar provas e questionários: Órgão implementador.

QUAL O PÚBLICO-ALVO A SER AVALIADO?

Tal como o termo é geralmente entendido, as avaliações nacionais referem-se a pesquisas realizadas nos sistemas educacionais. Essa conotação, porém, nem sempre existiu. Quando a primeira avaliação nacional foi realizada nos Estados Unidos (em 1969), tanto alunos que frequentavam escolas quanto grupos da população que não estavam na escola (jovens de 17 e 18 anos e adultos jovens de 26 a 35 anos) foram avaliados em cidadania, leitura e ciências. A avaliação da população fora da escola foi suspensa, contudo, em razão do custo elevado (Jones, 2003). Posteriormente, foram realizadas pesquisas de letramento de adultos, mas sem relação com as avaliações nacionais.

A questão da avaliação de crianças mais jovens fora da escola tem mais relevância em muitos países em desenvolvimento do que nos Estados Unidos, porque neles existem muitas crianças em idade escolar que não frequentam a escola. Obviamente, o aproveitamento (ou falta dele) dessas crianças é uma questão que interessa a formuladores de políticas e a políticos, e pode ter particular relevância para o setor de educação não formal. No entanto, é difícil imaginar sua inclusão numa avaliação nacional convencional. Embora determinados grupos de jovens fora da escola possam ser avaliados usando-se provas nacionais de avaliação num estudo separado, os métodos de avaliação e procedimentos de amostragem geralmente seriam muito diferentes, e as circunstâncias específicas dessas crianças (por exemplo, necessidades especiais, desvantagem socioeconômica ou distância da escola) teriam de ser levadas em conta.

No que se refere a crianças em idade escolar, os formuladores de políticas querem informações sobre seus conhecimentos e habilidades em determinados pontos de suas vidas escolares. É preciso decidir se os públicos-alvos serão definidos em função da idade ou da série, ou por uma combinação de ambas. Nos países em que existe grande variação na idade em que os alunos entram na escola, e se adota a política de não promoção, alunos de uma mesma faixa etária não estarão concentrados na mesma série. Nesse caso, existe uma sólida base para se preferir a segmentação por nível de ensino, e não por idade.

A série a ser avaliada normalmente deveria ser definida pelas necessidades de informações do Ministério da Educação. Se, por exemplo, o ministério está interessado em saber sobre os níveis de aproveitamento dos alunos que completaram o ensino básico, pode solicitar que uma avaliação nacional seja realizada no final do último ano (5ª ou 6ª série, em muitos países). O ministério também pode solicitar uma avaliação nacional da 3ª ou 4ª séries se precisar de dados sobre como os alunos estão se saindo na metade do ciclo do ensino básico. Essa informação poderia então ser utilizada para introduzir medidas corretivas (como cursos durante o trabalho para os professores), a fim de resolver problemas em aspectos específicos do currículo identificados na avaliação.

As séries em que são feitas as avaliações nacionais têm variado de país para país. Nos Estados Unidos, os níveis de aproveitamento dos alunos são avaliados nas 4ª, 8ª e 12ª séries; na Colômbia, o aproveitamento é avaliado nas 3ª, 5ª, 7ª e 9ª séries; no Uruguai, na pré-escola e nas 1ª, 2ª e 6ª séries; no Sri Lanka, nas 4ª, 8ª e 10ª séries. Nos países anglófonos da África, um consórcio regional de sistemas educacionais – Consórcio da África Austral e Oriental para Monitoramento da Qualidade da Educação (SACMEQ, na sigla em inglês) – avaliou alunos da 6ª série. Os países da África francófona que integram o Programa de Análise dos Sistemas Educacionais (PASEC, na sigla em francês) da Conferência dos Ministros da Educação dos Países de Língua Francesa (CONFEMEN, na sigla em francês) avaliaram alunos nas 2ª e 5ª séries.

Às vezes, a seleção da série é ditada por considerações pragmáticas. O Ministério Federal da Educação da Nigéria decidiu avaliar alunos na 4ª série porque, se a prova fosse feita em qualquer nível inferior, teria sido necessário traduzi-la para vários idiomas locais. E não se considerou adequado aplicar as provas em séries mais elevadas, porque tanto os alunos quanto os professores estariam voltados para os exames de admissão ao ensino fundamental.

Relativamente poucos países realizam avaliações em larga escala nas 1ª, 2ª ou 3ª séries. Os alunos nesses níveis talvez não consigam seguir instruções ou lidar com as tarefas cognitivas exigidas pela avaliação ou com o desafio de completar provas de múltipla escolha. Um estudo jamaica-

no observou que um número considerável de alunos da 1ª série não foi capaz de reconhecer as letras do alfabeto (Lockheed e Harris, 2005). No entanto, deve-se ter em mente que, como as informações sobre os padrões de aprendizado dos alunos nas séries iniciais podem ser cruciais para os processos de reforma, devem existir procedimentos alternativos para monitorar esses padrões.

Responsabilidade de selecionar o público-alvo a ser avaliado: Ministério da Educação e CCN.

A AVALIAÇÃO COBRIRÁ TODO O PÚBLICO-ALVO OU APENAS UMA AMOSTRA?

A maioria dos estudos nacionais, bem como todos os regionais e internacionais, usa abordagens baseadas em amostras para determinar os níveis nacionais de aproveitamento. Algumas avaliações nacionais têm usado tanto o censo quanto amostragem (por exemplo, Costa Rica, Cuba, França, Honduras, Jordânia, México e Uruguai), enquanto a maior parte das avaliações subnacionais coleta dados do censo (por exemplo, Minas Gerais, Paraná e São Paulo, no Brasil; Bogotá, na Colômbia, e Aguascalientes, no México) (veja Crespo, Soares e Mello e Souza, 2000). Vários fatores favorecem a utilização de uma amostra se o objetivo é obter informações sobre o funcionamento do sistema educativo como um todo para fins de políticas. Esses fatores incluem: menores custos na aplicação da prova e na limpeza e gestão dos dados; menos tempo para análise e relatórios; maior exatidão, devido à possibilidade de assegurar supervisão mais intensa do trabalho de campo e de preparação dos dados (Ross, 1987).

Como observado no Capítulo 3, o objetivo da avaliação é fundamental para determinar quando se deve testar uma amostra ou todos os alunos de uma população-alvo. Por um lado, a decisão de envolver toda a população-alvo pode refletir a intenção de partilhar com as escolas, os professores ou mesmo os alunos a responsabilidade pelo aprendizado. Isso facilita o uso de sanções (incentivos ou penalidades), o fornecimen-

to de informações às escolas sobre seu aproveitamento individual e a publicação de tabelas classificativas, bem como a identificação das escolas com maior necessidade de assistência (como feito no Chile e no México, por exemplo).

Por outro lado, a abordagem baseada em amostras permitirá a detecção de problemas apenas no nível do sistema como um todo. Não será possível identificar as escolas específicas que estejam necessitando de apoio, embora se possam identificar os tipos ou categorias de escolas (por exemplo, pequenas escolas rurais) que requeiram atenção. Essa abordagem também é capaz de identificar problemas relativos à equidade de gênero ou de etnia. Um argumento contra o uso de uma abordagem baseada em amostras é que, como a avaliação não tem fortes interesses associados ao aproveitamento, alguns alunos não estarão motivados a levar a prova a sério.

No entanto, esse não foi o caso em muitos países – inclusive na África do Sul – em que alguns alunos receavam que seu desempenho nos testes do TIMMS (Tendências no Estudo Internacional de Matemática e Ciências) fosse contar para seus resultados escolares oficiais. É interessante notar que houve trapaça durante a aplicação da prova, provavelmente por causa da percepção de que interesses relativamente elevados estavam ligados ao desempenho (veja A.4 no Apêndice A).

As vantagens e desvantagens da utilização de uma avaliação nacional para promover a noção de responsabilidade das escolas, dos professores e dos alunos estão contidas na Tabela 4.2. Os tópicos listados derivam, na maior parte dos casos, de estudos sobre os efeitos de exames públicos nos quais existiam fortes interesses em jogo, não de um estudo de avaliações nacionais. No entanto, devem ser relevantes para as avaliações nacionais baseadas em censos, pelo menos para aquelas que funcionam como substituto de exames públicos (como nos Estados Unidos e em alguns países da América Latina).

Responsabilidade de decidir sobre o uso de uma amostra ou do censo: Ministério da Educação

TABELA 4.2

Vantagens e Desvantagens da Avaliação Baseada no Censo para Manter a Responsabilidade das Escolas

| Vantagens | Desvantagens |
|--|---|
| Incide sobre o que são considerados aspectos importantes da educação. | Tende a conduzir à negligência das áreas de estudo que não são testadas. |
| Destaca aspectos importantes de matérias específicas. | Tende a conduzir à negligência de aspectos dos temas não testados (como fluência oral na língua). |
| Ajuda a garantir que os alunos alcancem um padrão aceitável antes da promoção. | Tem contribuído para a evasão precoce e a repetência. |
| Permite comparações diretas entre as escolas. | Resulta na classificação injusta de escolas em que diferentes contextos sociais são atendidos e os resultados não são significativamente diferentes. |
| Cria confiança pública no desempenho do sistema. | Tem levado a fraudes durante a aplicação do teste e à subsequente correção de resultados. |
| Exerce pressão sobre os alunos para que aprendam. | Tende a enfatizar a memorização e a aprendizagem por repetição. |
| Os resultados de algumas escolas e alunos elevam os níveis de desempenho do teste. | Melhor desempenho pode estar limitado a um teste específico e não será evidente em outros testes da mesma matéria. |
| Permite que os pais avaliem a eficácia de escolas e professores específicos. | Conduz a uma avaliação injusta da eficácia com base na pontuação no teste, e não leva em conta outros fatores reconhecidamente associados ao aproveitamento do aprendizado. |
| Tende a ser popular com os políticos e os meios de comunicação. | Raramente responsabiliza os políticos por não darem apoio à provisão de recursos educacionais. |

Fonte: Autores.

O QUE SERÁ AVALIADO?

Todas as avaliações nacionais medem os resultados cognitivos da instrução ou as habilidades acadêmicas nas áreas de letramento em linguagem e letramento em matemática, e isso reflete a importância desses resultados para a educação básica. Em alguns países, o conhecimento de ciências e de estudos sociais está incluído na avaliação. Seja qual for o âmbito da avaliação, a explicitação de um marco de referência, ou marco con-

textual, é importante em duas etapas: na construção dos instrumentos de avaliação e, depois, durante a interpretação dos resultados. O marco de referência pode estar disponível em um documento de currículo se, por exemplo, o documento apresenta expectativas de aprendizagem claramente priorizadas e postas em prática. Na maioria dos casos, no entanto, tal marco de referência não estará disponível, e os encarregados da avaliação nacional terão de construí-lo. Para realizar essa tarefa, será necessária uma cooperação estreita entre o órgão de avaliação, os responsáveis pelos currículos e outras partes interessadas.

Os marcos de referência de uma avaliação tentam esclarecer em detalhes, numa avaliação em larga escala, o que está sendo avaliado, como está sendo avaliado e por que está sendo avaliado (veja Kirsch, 2001). O objetivo desse marco é dar transparência ao processo de avaliação e aos pressupostos que o informam, não apenas para os elaboradores dos testes, mas também para um público muito maior, incluindo professores, pessoal de currículo e formuladores de políticas. O marco de referência geralmente começa com uma definição geral ou uma declaração de propósito que norteia o fundamento lógico da avaliação e especifica o que deve ser medido em termos de conhecimentos, habilidades e outros atributos. Em seguida, identifica e descreve os vários desempenhos ou comportamentos que revelarão esses atributos, identificando um número específico de tarefas ou variáveis características a serem utilizadas no desenvolvimento da avaliação e indicando como esses desempenhos devem ser utilizados para avaliar o aproveitamento dos alunos (Mullis *et al.*, 2006).

Muitas avaliações nacionais se basearam em uma análise de conteúdo, em determinado nível de ensino, daquilo que se supõe que os alunos devam ter aprendido como resultado da exposição a um currículo prescrito ou pretendido. Normalmente, essa análise é feita em uma matriz que tem os comportamentos cognitivos no eixo horizontal, e os temas ou áreas de conteúdo no eixo vertical. Assim, a interseção de um comportamento cognitivo e de uma área de conteúdo representará um objetivo do aprendizado. As células podem ser ponderadas em termos de importância.

Avaliações nacionais (e internacionais) recentes basearam-se em pesquisas sobre o desenvolvimento alcançado pelos alunos quanto a obras literárias e habilidades matemáticas que podem estar ou não represen-

tadas nos currículos nacionais. Por exemplo, no documento Marco de Referência e Especificações para o Estudo Internacional sobre o Progresso do Letramento em Leitura (PIRLS, na sigla em inglês), realizado em 2006 pela Associação Internacional para Avaliação do Aproveitamento Escolar (IEA, na sigla em inglês), o letramento em leitura é definido como “a capacidade de compreender e utilizar as formas da linguagem escrita exigidas pela sociedade e/ou valorizadas pelo indivíduo. Leitores jovens podem construir significado a partir de uma variedade de textos. Eles leem para aprender, participar de comunidades de leitores na escola e na vida cotidiana, e por prazer” (Mullis *et al.*, 2006, 3). A partir dessa definição, é evidente que ler é muito mais do que decodificar textos ou compreender o significado de uma passagem ou de um poema. O PIRLS aclarou ainda mais o que pretendia medir ao indicar o processo e as tarefas que seriam objetos da avaliação e os percentuais de itens da prova dedicados a cada um (Tabela 4.3).

TABELA 4.3
Processos de Compreensão de Leitura do PIRLS

| Processos de compreensão | Exemplos de tarefas | Itens |
|---|--|-------|
| Focalizar e recuperar informação explícita | Procurar ideias específicas. Encontrar definições ou frases. Identificar o cenário de uma história (por exemplo, tempo, lugar). Encontrar a tópicos frasal ou ideia principal (expressa explicitamente). | 20% |
| Fazer inferências simples | Inferir que um acontecimento causou outro. Identificar generalizações no texto. Descrever a relação entre os personagens. Determinar o referente de um pronome. | 30% |
| Interpretar e integrar ideias e informações | Determinar a mensagem geral ou tema. Contrastar as informações do texto. Inferir o sentido de humor ou o tom de uma história. Encontrar uma aplicação na vida real para a informação contida no texto. | 30% |
| Examinar e avaliar conteúdo, linguagem e elementos textuais | Avaliar a probabilidade de que os eventos descritos possam acontecer. Descrever como o autor concebeu um final inesperado. Julgar a integridade ou a clareza da informação no texto. Determinar as perspectivas do autor. | 20% |

Fonte: Campbell *et al.*, 2001; Mullis *et al.*, 2006.

O documento que define o marco de referência especificou que a avaliação usaria cadernos de prova com cinco passagens literárias e cinco informativas, e que cada passagem seria seguida de 12 questões, sendo metade de múltipla escolha e as demais de respostas fechadas. Além disso, indicou que, como as atitudes e os comportamentos relativos à leitura são importantes para o desenvolvimento do hábito da leitura por toda a vida e estão associados ao aproveitamento em leitura, o PIRLS incluiria no questionário alguns itens destinados a avaliar as atitudes e os comportamentos dos alunos com relação à leitura. A seleção de alunos que se encontravam no quarto ano de escolaridade como a população-alvo para a avaliação foi justificada com base no fato de que essa etapa representava a fase de transição entre aprender a ler e ler para aprender.

No seu marco de referência para a avaliação, o PIRLS reconheceu os dois principais propósitos da leitura para os alunos:

- Ler para ganhar experiência de leitura
- Ler para adquirir e usar informação.

O documento também ofereceu uma justificativa detalhada para a ênfase dada pelo PIRLS à obtenção de informações adicionais sobre o ambiente e o contexto em que os alunos aprendem a ler. Essa ênfase levou à inclusão de itens no questionário para identificar as características da família que podem incentivar as crianças a aprender a ler: as atividades dos pais relacionadas ao letramento, a língua falada em casa, as ligações entre a casa e a escola, e atividades de letramento fora da escola. Os itens relativos à escola abrangiam recursos da escola que podem afetar direta ou indiretamente o aproveitamento da leitura. O documento-quadro também justificava a avaliação de variáveis relativas à sala de aula, tais como abordagens instrucionais e natureza da formação de professores.

Outro modo de basear um instrumento de avaliação nas expectativas ou prescrições contidas no currículo – o que é viável no caso de alunos mais velhos – é construir um teste que reflita o conhecimento e as habilidades de que os alunos possivelmente precisarão e que poderão desenvolver na vida adulta. O Programa Internacional de Avaliação de Alunos (PISA) forneceu um exemplo desse método quando decidiu

avaliar o “letramento matemático” de alunos de 15 anos, definido como a “capacidade de identificar e compreender o papel que a matemática desempenha no mundo, de tomar decisões bem fundamentadas e usar a matemática para participar de trabalhos que atendam às necessidades da vida do indivíduo como um cidadão construtivo, interessado e reflexivo” (OCDE, 2003, 24) (veja B.3 no Apêndice B). Embora essa abordagem se enquadre bem num estudo internacional, já que a alternativa de conceber um instrumento de avaliação igualmente apropriado para uma variedade de currículos é obviamente problemática, também pode ser utilizada numa avaliação nacional.

Algumas poucas avaliações nacionais coletaram informações sobre os resultados afetivos (por exemplo, as atitudes dos alunos com relação à escola e sua autoestima). Na Colômbia, são avaliadas as atitudes dos alunos relativas à paz. Embora esses resultados sejam muito importantes, sua medição tende a ser menos confiável que a medição de resultados cognitivos, e as análises com base neles têm-se revelado de difícil interpretação. No Chile, as dificuldades técnicas para medir valores e atitudes dos alunos com relação ao aprendizado levaram ao abandono dessas áreas (veja A.7 no Apêndice A).

Uma avaliação em larga escala (Monitoramento do Aproveitamento do Aprendizado) avaliou as “habilidades para a vida”, definidas como o conhecimento e as atitudes dos alunos sobre saúde e nutrição, meio ambiente, responsabilidade cívica, e ciência e tecnologia (Chinapah, 1997). Embora seja geralmente aceito que as habilidades para a vida sejam importantes e devam ser ensinadas, há bastante discordância sobre sua exata natureza. Sua medição também tem se revelado difícil.

A maioria das avaliações nacionais coleta informações sobre os alunos, a escola e fatores domésticos considerados relevantes para o aproveitamento (por exemplo, gênero dos alunos; sua vida escolar, incluindo repetência; recursos nas escolas, incluindo disponibilidade de livros didáticos, nível de formação de professores e suas qualificações; e situação socioeconômica das famílias). As informações são, em geral, obtidas por meio de questionários (e, ocasionalmente, em entrevistas) aplicados aos alunos, professores, diretores e, às vezes, aos pais, ao mesmo tempo em que estão sendo aplicados os instrumentos de avaliação.

A identificação de fatores contextuais relacionados com o aproveitamento dos alunos pode ajudar a identificar as variáveis manipuláveis, ou seja, os fatores que podem ser alterados pelos formuladores de políticas, como a regulamentação do tempo alocado para as áreas de currículo, a oferta de livros didáticos e o tamanho da turma. No entanto, os dados contextuais coletados em alguns estudos nacionais (e internacionais) não podem desempenhar esse papel, porque não medem adequadamente as condições em que vivem os alunos. A determinação da situação econômica, por exemplo, pode estar baseada numa lista de itens que inclui um carro, uma televisão e uma torneira – embora isso não se ajuste à situação de um país em que a maioria da população vive, pelo menos durante parte do ano, com o equivalente a menos de US\$1 por dia. Além disso, apesar da relevância de fatores como estado de saúde e estado nutricional, não há como obter informações sobre eles (Naumann, 2005).

Em algumas avaliações, o aproveitamento dos professores também foi avaliado. No Vietnã (veja A.2 no Apêndice A) e em vários países africanos que participaram dos estudos do SACMEQ (veja C.1 do Apêndice C), os professores tiveram de responder aos mesmos itens do teste aplicado a seus alunos, como forma de se obter uma percepção de seus níveis de domínio do assunto. Em Uganda, foi possível saber o quanto os professores afirmavam estar familiarizados com os principais documentos curriculares oficiais.

Responsabilidade de decidir o que será avaliado: Ministério da Educação, CCN, com insumos fornecidos pelo órgão implementador.

COMO SERÁ AVALIADO O APROVEITAMENTO?

Deve-se conceber um instrumento, ou mais de um, para fornecer a informação que a avaliação nacional pretende obter. Assim como variam as finalidades das avaliações nacionais e seus usos propostos, variam também os instrumentos utilizados nas avaliações e as formas como os resultados são apresentados.

Algumas avaliações nacionais apresentam resultados em função das características da distribuição das pontuações nas provas – por exemplo, a média percentual de itens que os alunos responderam corretamente e como as pontuações se distribuíram em torno da média. Ou os resultados podem ser apresentados numa escala relativa a uma média arbitrária (como 500, por exemplo) e a um desvio-padrão também arbitrário (como 100, por exemplo). Embora essas pontuações possam ser usadas para comparar o desempenho de subgrupos na amostra, têm utilização limitada numa avaliação nacional, principalmente porque pouco nos dizem a respeito do nível de conhecimentos dos alunos sobre as matérias ou das habilidades reais que adquiriram.

Para resolver essa questão e tornar os resultados de uma avaliação mais significativos para as partes interessadas, um número crescente de avaliações nacionais procura apresentar os resultados de uma forma que especifica o que os alunos sabem e o que não sabem e identifica os pontos fortes e fracos em seus conhecimentos e habilidades. Essa abordagem envolve a combinação das pontuações dos alunos com as descrições das tarefas que são capazes de fazer (por exemplo, “pode ler em determinado nível de compreensão” ou “pode realizar operações matemáticas básicas”). Os desempenhos podem ser classificados de diversos modos (por exemplo, “satisfatório” ou “insatisfatório”, “básico”, “proficiente” ou “avançado”), e pode-se determinar a proporção de alunos em cada nível. Estabelecer a correspondência entre as pontuações dos alunos e os níveis de aproveitamento é uma tarefa complexa que envolve o julgamento de especialistas em currículos e em análise estatística.

A forma como os resultados serão descritos deve ser decidida na fase de elaboração das provas. Assim, a elaboração poderia começar com a especificação de um quadro de referência no qual são explicitadas as expectativas a respeito do aprendizado; em seguida, seriam escritos os itens do teste destinados a avaliar em que medida os alunos atendem às expectativas. No entanto, se os itens não atenderem a certos critérios na fase do pré-teste, incluindo a medida em que conseguem captar as diferenças entre os alunos, não poderão ser usados no instrumento de avaliação final. Devem ser tomados cuidados para garantir que objetivos

importantes do currículo estejam representados numa avaliação, mesmo que nenhum aluno no pré-teste forneça evidência de alcançá-los.

A maioria das avaliações nacionais e internacionais utiliza, em grande medida, o formato de múltipla escolha em seus instrumentos. Muitas vezes, aqueles itens serão complementados por perguntas abertas que requerem que o aluno escreva uma palavra, frase ou sentença. Exemplos de itens de múltipla escolha e de perguntas abertas são fornecidos nos Quadros 4.2 e 4.3, respectivamente.

Em várias avaliações nacionais (por exemplo, a NAEP, nos Estados Unidos, e a Avaliação Nacional de Leitura em Inglês, na Irlanda) e internacionais (por exemplo, o TIMSS e o PISA), cada aluno responde a apenas uma fração do número total de itens utilizados numa avaliação (veja A.8 no Apêndice A; B.1 e B.3 no Apêndice B). Essa abordagem amplia as áreas do currículo cobertas pela prova sem colocar um fardo muito grande sobre os alunos individualmente. Também permite o uso de passagens dissertativas (por exemplo, uma crônica ou um artigo de jornal) na avaliação da compreensão em leitura. Em outras avaliações, todos os alunos respondem ao mesmo conjunto de itens. Embora haja algumas vantagens em haver alunos individuais respondendo a apenas uma fração de itens, também existem desvantagens, especialmente para os países que estão começando um programa de avaliação nacional. A administração (por exemplo, impressão e distribuição) é mais complexa, assim como a pontuação e o ordenamento das pontuações, e análises envolvendo dados individuais de alunos podem ser problemáticas (veja Sofroniou e Kellaghan, 2004).

A questão do idioma de avaliação geralmente recebe menos atenção do que merece. Ela está associada a dois problemas. Em primeiro lugar, embora em muitos países existam grandes grupos minoritários (e, às vezes, majoritários) para os quais a língua de instrução não é seu idioma materno, os alunos são normalmente avaliados na língua de instrução. Em Uganda, por exemplo, a maioria dos alunos faz as provas em sua segunda língua (veja A.9 no Apêndice A). O baixo desempenho em provas é atribuído a essa prática, bem como o são os geralmente limitados progressos acadêmicos dos alunos e o abandono precoce da escola (Naumann, 2005).

QUADRO 4.2

Exemplos de Itens de Múltipla Escolha

Matéria: Geografia

O rio Volga fica na:

- A. China
- B. Alemanha
- C. Federação Russa
- D. Suécia.

Matéria: Matemática

Uma foca tem de respirar enquanto dorme. Martin observou uma foca durante uma hora. No início da observação, a foca mergulhou até o fundo do mar e começou a dormir. Em oito minutos, flutuou lentamente até a superfície e respirou. Passados três minutos, ela estava de volta ao fundo do mar, e o processo recomeçou de forma muito regular. Depois de uma hora, a foca estava:

- A. no fundo da água
- B. a caminho da superfície
- C. respirando
- D. a caminho do fundo.

Fonte: Exemplo de matemática: OCDE, 2007. Reproduzido com permissão.

QUADRO 4.3

Exemplos de Itens Abertos

Matéria: Linguagem

ALTO é o oposto de BAIXO

Qual é o oposto de:

RÁPIDO _____ ESCURO _____

PESADO _____ VELHO _____

Matéria: Matemática

Use sua régua para desenhar um retângulo com um perímetro de 20cm. Identifique com o nome a largura e o comprimento.

Um segundo problema relativo à linguagem surge quando os instrumentos de avaliação precisam ser traduzidos em uma ou mais línguas. Quando se deseja fazer comparações entre desempenhos avaliados em diferentes línguas, a análise deve levar em conta a possibilidade de que eventuais diferenças possam ser atribuídas a diferentes graus de dificuldade linguística encontrados pelos alunos nas provas de avaliação. A questão é parcialmente resolvida com a mudança de palavras. Por exemplo, em uma avaliação internacional realizada em Gana, a palavra “neve” foi substituída por “chuva”. Se variações nas diferenças de idiomas estiverem associadas a variações nos fatores culturais e econômicos, o problema é agravado, pois pode ser difícil assegurar a equivalência, em todos os idiomas em que é redigido um teste, entre a forma como são elaboradas as perguntas e a adequação cultural dos conteúdos. Por exemplo, o material apropriado ao contexto de alunos nas áreas rurais – cobrindo caça, mercado ou feira, atividades agrícolas e jogos locais – pode ser desconhecido para os alunos em áreas urbanas.

Quaisquer que sejam os detalhes do método utilizado, a avaliação precisa fornecer informações válidas e confiáveis. A validade tem diversas facetas, incluindo a adequação de um instrumento de avaliação para representar o conceito (por exemplo, letramento em leitura) ou a área de currículo (por exemplo, estudos sociais) identificados no contexto de avaliação. O julgamento de especialistas em currículo é importante aqui. Além disso, o instrumento de avaliação deve medir apenas o que se pretende medir com ele. Assim, um teste de matemática ou ciências deve avaliar os conhecimentos e habilidades dos alunos nessas áreas, e não sua competência em linguagem. A confiabilidade dos procedimentos de avaliação em avaliações nacionais geralmente envolve saber estimar em que medida os itens individuais de um teste avaliam o conceito geral que o teste pretende medir e, no caso de itens em aberto, a medida em que dois ou mais membros da equipe concordam em suas pontuações.

Responsabilidade de decidir como será avaliado o aproveitamento: Órgão de implementação.

COM QUE FREQUÊNCIA SERÃO REALIZADAS AS AVALIAÇÕES?

A frequência com que se realiza uma avaliação nacional varia de país para país, podendo ir de avaliações anuais até uma a cada 10 anos. Pode existir a tentação de avaliar o aproveitamento na mesma área de currículo e em um mesmo público-alvo todos os anos, mas essa frequência é desnecessária, bem como muito dispendiosa, se o objetivo for monitorar os padrões nacionais. Nos Estados Unidos, leitura e matemática são avaliadas a cada dois anos, e outras matérias têm avaliações menos frequentes. A avaliação internacional do letramento em leitura (PIRLS) teve um espaço de cinco anos entre a primeira aplicação e a segunda (2001-06). No Japão, o aproveitamento nas áreas principais do currículo é avaliado a cada 10 anos, a fim de orientar a revisão de currículos e de livros didáticos (Ishino, 1995).

Se o objetivo da avaliação é manter os professores e as escolas – e até mesmo os alunos – responsáveis pelo aprendizado, as provas podem ser realizadas a cada ano. Além disso, dado que tal avaliação incide sobre o desempenho de indivíduos, bem como sobre o desempenho do sistema, todos (ou quase todos) os alunos no sistema educacional serão avaliados. Esse esquema foi adotado no Chile e na Inglaterra.

No entanto, se a finalidade da avaliação é apenas fornecer informações sobre o aproveitamento do sistema como um todo, talvez seja adequado fazer a avaliação de uma amostra de alunos em determinada área de currículo a cada três a cinco anos. Como os sistemas educacionais não mudam rapidamente, avaliações mais frequentes provavelmente não registrariam mudanças. É mais que provável que avaliações feitas com frequência excessiva limitem o impacto dos resultados e resultem em custos desnecessários.

Responsabilidade por decidir a frequência da avaliação: Ministério da Educação

COMO DEVE SER DIVULGADO O APROVEITAMENTO DOS ALUNOS?

Embora os formuladores de políticas provavelmente prefiram estatísticas resumidas, a evidência sobre a multidimensionalidade do aprovei-

tamento sugere que um único índice de aproveitamento, tal como a pontuação total de uma prova, pode ocultar informações importantes. Uma abordagem alternativa é fornecer informações diferenciadas que reflitam os pontos fortes e fracos no currículo de um país. A informação seria ainda mais valiosa se distinguisse entre o conhecimento dos alunos sobre fatos básicos e habilidades básicas e sua compreensão mais profunda, ou de ordem superior.

Para descrever o aproveitamento dos alunos em avaliações nacionais, tem-se usado uma variedade de procedimentos, e isso reflete a riqueza de dados que uma avaliação pode fornecer (veja o Volume 5, *O uso dos resultados da avaliação do desempenho educacional*). A seleção de um ou mais procedimentos deve ser orientada pelas necessidades de informação do Ministério da Educação e de outros interessados.

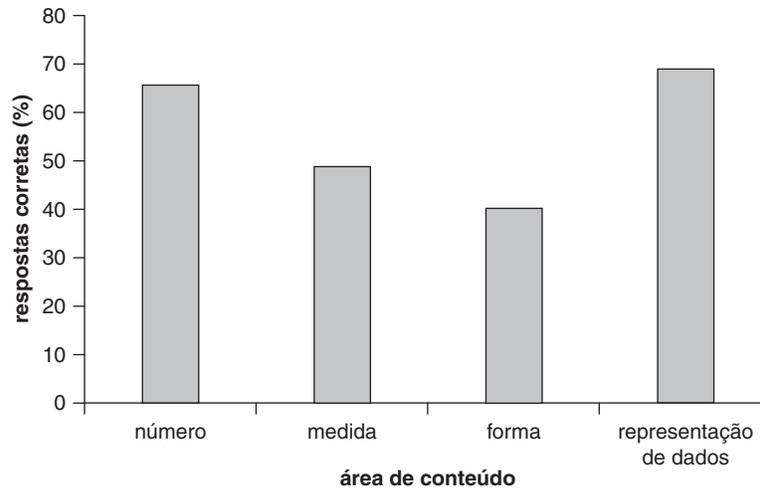
Informações no Nível de Item

Esta informação envolve pouco mais além de simplesmente relatar o percentual de alunos que responderam corretamente a itens individuais. Uma avaliação nacional pode revelar que a maioria dos alunos teve desempenho insatisfatório em uma questão de matemática que envolvia o uso de índices, ou que praticamente todos os alunos foram capazes de estabelecer associações entre palavras simples e imagens. Em Gana, por exemplo, apenas 1% dos alunos respondeu corretamente a uma pergunta do TIMSS sobre a refração da luz (Gana, Ministério da Educação, Juventude e Esportes, 2004). Esse tipo de informação, embora muito detalhada para a elaboração de políticas nacionais, provavelmente interessaria aos que trabalham com currículos, aos formadores de professores e, possivelmente, a autores de livros didáticos.

Desempenho em Domínios de Currículo

Os itens podem ser agrupados em unidades curriculares ou domínios, e as pontuações das provas podem ser divulgadas em função do desempe-

FIGURA 4.1

Percentual Médio de Respostas Corretas em Matemática, por Áreas de Conteúdo no Lesoto

Fonte: Lesoto, Conselho de Testes do Lesoto e Centro Nacional de Desenvolvimento de Currículos, 2006.

nho em cada domínio. Itens de leitura, por exemplo, têm sido classificados por capacidade de recuperar informações de um texto, fazer inferências a partir de um texto, interpretar e integrar informações, examinar e avaliar as informações contidas no texto (Eivers *et al.*, 2005). A Figura 4.1 ilustra como o Lesoto divulgou o desempenho em matemática por área de conteúdo.

Padrões de Desempenho

O desempenho em avaliações nacionais e internacionais descreve o que se requer dos alunos nas provas para que consigam uma classificação no nível “básico”, “proficiente” ou “avançado” numa área de currículo. O número de níveis pode variar (veja A.2 no Apêndice A para uma descrição de seis níveis de proficiência de leitura utilizados numa avaliação nacional no Vietnã, e C.1 no Apêndice C para oito níveis de leitura e oito

níveis de habilidade matemática utilizados no SACMEQ). A seleção dos pontos de corte que diferenciarão os níveis envolve o uso de dados estatísticos e julgamentos subjetivos.

Padrões de Domínio

Os níveis de domínio podem basear-se na pontuação geral obtida na prova (por exemplo, respostas corretas a determinado percentual de itens de teste). No Sri Lanka, o nível de domínio para uma avaliação nacional da 4ª série foi fixado em 80%. Menos de 40% dos alunos atingiram esse nível no teste de seu idioma materno ou em matemática, e menos de 10% o alcançaram no teste de inglês (Perera *et al.*, 2004). Os níveis de domínio também podem se basear no atingimento de determinado nível de desempenho. Nos Estados Unidos, o estado de Connecticut usa cinco níveis de desempenho (“abaixo do básico”, “básico”, “proficiente”, “conforme a meta” e “avançado”). A “meta” é definida como um nível desafiador, embora fosse razoavelmente possível esperá-la dos alunos, e é aceita como o nível de domínio. Os dados da Tabela 4.4 mostram que bem mais da metade dos alunos nas 3ª e 4ª séries alcançou o nível “conforme a meta” (ou “de domínio”) em todas as três áreas de currículo.

Responsabilidade por decidir como o aproveitamento dos alunos é apresentado: Órgão de implementação, com insumos da CCN.

QUE TIPOS DE ANÁLISES ESTATÍSTICAS DEVEM SER FEITAS?

Algumas análises serão ditadas pelas questões de políticas que inicialmente motivaram a avaliação. A maioria das avaliações nacionais fornece evidências sobre o aproveitamento por gênero, região, área urbana ou rural, etnia ou língua, e tipo de escola frequentada (pública ou privada). Algumas avaliações também fornecem dados sobre a qualidade das instalações escolares (por exemplo, no Quênia). As análises que envolvem essas variáveis são relativamente simples e têm significado óbvio para

TABELA 4.4
Percentual que Alcançou a Meta ou o Nível de Domínio por Série, Connecticut, 2006

| Série | Matemática | | Leitura | | Escrita | |
|----------------|----------------------|-----------------------|----------------------|-----------------------|----------------------|-----------------------|
| | Na meta ou acima (%) | Avançado ou acima (%) | Na meta ou acima (%) | Avançado ou acima (%) | Na meta ou acima (%) | Avançado ou acima (%) |
| 3 ^a | 56 | 22 | 54 | 17 | 61 | 22 |
| 4 ^a | 59 | 22 | 58 | 16 | 63 | 22 |

Fonte: Departamento de Educação de Connecticut, 2006.

os formuladores de políticas e os políticos. No entanto, não representam adequadamente a complexidade dos dados. São necessárias formas mais complexas de análise se quisermos, por exemplo, lançar luz sobre os fatores relativos à escola e sobre os antecedentes e contextos que contribuem para o aproveitamento. Exemplos do uso de procedimentos estatísticos complexos encontram-se na descrição da avaliação nacional do Vietnã (veja A.2 no Apêndice A).

Devem-se levar em conta as limitações das análises e os problemas de se inferir causalidade a partir de estudos que coletam, ao mesmo tempo, dados sobre aproveitamento e sobre outras variáveis. Embora seja difícil, e às vezes impossível, separar os efeitos da comunidade, da família e de fatores escolares sobre o aprendizado dos alunos, essa complexidade não tem impedido que algumas pesquisas façam interpretações causais dos dados obtidos em avaliações nacionais e internacionais.

Responsabilidade de decidir sobre os métodos de análise estatística: Órgão de implementação.

COMO DEVEM SER APRESENTADOS E USADOS OS RESULTADOS DE UMA AVALIAÇÃO NACIONAL?

Caso se deseje que os resultados de uma avaliação nacional afetem a política nacional de educação, esses devem ser divulgados o mais rapidamente possível após a conclusão da análise dos dados. No passado, os relatórios técnicos que continham uma quantidade considerável de dados tendiam a ser a única forma de apresentação dos resultados. Alguns grupos de usuários (por exemplo, os professores no Chile; ver A.7 no Apêndice A), contudo, consideraram que a linguagem desses relatórios era excessivamente técnica. Como resultado, cresce a cada dia o reconhecimento de que é indispensável adotar outras formas de divulgação. As alternativas incluem relatórios resumidos centrados nos principais resultados e especialmente preparados para os formuladores de políticas (que dispõem de pouco tempo para ler textos longos); comunicados de imprensa; relatórios especiais para rádio e televisão; e relatórios separados para escolas, professores, elaboradores de

currículos e formadores de professores. Em alguns países (por exemplo, Sri Lanka), são preparados relatórios distintos para cada província. Um relatório na Etiópia foi traduzido para quatro línguas principais. As necessidades de informação das partes interessadas devem determinar o conteúdo de relatórios adicionais.

Ainda na fase de planejamento, o Ministério da Educação deve fazer a dotação orçamentária adequada para a elaboração do relatório e sua divulgação. Em colaboração com a comissão de coordenação nacional, deve definir procedimentos para comunicar às partes interessadas os resultados das avaliações nacionais. As estratégias apropriadas para comunicar os resultados devem levar em conta o fato de que, entre os usuários (quer se trate de administradores ou professores), há grande variação na capacidade de compreender a informação estatística e aplicá-la na tomada de decisões. Obviamente, não faz sentido produzir relatórios se as informações ali contidas não forem adequadamente divulgadas. Assim, também é necessário definir uma estratégia de divulgação para que a informação relevante atinja todas as partes interessadas. A estratégia deve identificar potenciais usuários (principais instituições e indivíduos) e seus níveis de competência técnica.

Os resultados da avaliação nacional têm sido utilizados para estabelecer padrões de referência que permitam acompanhar os níveis de aproveitamento do aprendizado (por exemplo, no Lesoto), reformar currículos, fornecer dados de linha de base sobre a quantidade e qualidade de materiais educacionais nas escolas (por exemplo, no Vietnã), identificar correlatos de aproveitamento e diagnosticar aspectos do currículo que não estão sendo dominados pelos alunos. O Uruguai, por exemplo, usou os resultados de sua avaliação nacional a fim de preparar manuais para professores e identificar conteúdos curriculares e áreas comportamentais que, posteriormente, ajudaram a direcionar um programa em grande escala para a capacitação durante o trabalho (veja A.3 no Apêndice A).

O Volume 5 contém uma extensa parte sobre a elaboração de relatórios e a utilização dos resultados da avaliação nacional.

Responsabilidade de apresentar e utilizar os resultados da avaliação nacional: Órgão implementador, Ministério da Educação, CCN, formadores de professores, autoridade curricular, professores.

QUAIS SÃO OS COMPONENTES DE CUSTO DE UMA AVALIAÇÃO NACIONAL?

O custo de uma avaliação nacional vai variar muito entre os países, dependendo dos níveis salariais do pessoal e dos custos dos serviços. Dentro de um país, o custo também varia em função de alguns dos seguintes fatores, ou de todos (Ilon, 1996):

- *Órgão implementador.* Os custos variarão dependendo de o órgão ter as instalações e os conhecimentos necessários ou de precisar, respectivamente, aprimorá-las ou contratar consultores em tempo integral ou parcial. O custo de fornecer instalações e equipamentos, incluindo computadores e software, também precisa ser levado em conta.
- *Conteúdo e elaboração do instrumento.* As opções para a seleção do conteúdo e da forma da avaliação devem levar em conta o custo e outros fatores como validade e facilidade de aplicação. A construção de itens de múltipla escolha é mais cara que a de itens em aberto, mas, em geral, o trabalho de pontuação da múltipla escolha é menos dispendioso. O custo da tradução de provas, questionários e manuais, bem como de treinar os redatores de itens, também precisa ser considerado.
- *Número de escolas e alunos participantes.* Uma avaliação baseada no censo será obviamente mais cara que outra baseada numa amostra. Os custos aumentam se houver necessidade de dados confiáveis para setores do sistema (estados ou municípios, por exemplo). Concentrar a avaliação numa faixa de idade específica tenderá a ser mais caro que focalizar um nível de ensino, porque alunos de todas as idades podem estar espalhados em várias séries, e isso exigiria material de avaliação complementar e sessões adicionais de provas.
- *Administração.* A coleta de dados tende a ser o componente mais caro de uma avaliação nacional. Ela envolve obter informações das escolas antes da avaliação; conceber, imprimir, embalar e enviar material de teste e questionários; e criar um sistema para aplicar os instrumentos. Os fatores que contribuem para o custo global incluem: número de escolas e alunos participantes; viagens; dificuldade de acesso às

escolas; alojamento para os entrevistadores (se necessário); coleta e retorno de questionários e provas concluídos.

- *Pontuação, gerenciamento de dados e entrada de dados.* Os custos variarão de acordo com alguns aspectos: o número de escolas, alunos, professores e pais que participarem; o número de itens em aberto; se os itens são pontuados manualmente ou a máquina; o número de estudos de confiabilidade feitos entre avaliadores; a qualidade da aplicação da prova e da pontuação.
- *Análise.* Os custos da análise dependerão do tipo de procedimentos de avaliação utilizados e da disponibilidade de tecnologia para pontuação e análise. Apesar de a pontuação a máquina ser normalmente considerada mais barata que a manual, esse fator de redução do custo pode não se aplicar a um país em que os custos de tecnologia são elevados e os custos trabalhistas, baixos.
- *Apresentação dos resultados.* A avaliação dos custos deve levar em conta o fato de que serão necessárias diferentes versões de um relatório para os formuladores de políticas, os professores e o público em geral; também se devem considerar a natureza e a amplitude da estratégia de divulgação do relatório.
- *Atividades de acompanhamento posterior.* Talvez haja necessidade de dotação orçamentária para atividades como treinamento dos professores durante o trabalho (em consequência dos resultados da avaliação nacional), sessões informativas para órgãos de currículos e análises secundárias dos dados. Também podem ser necessárias medidas para contornar a escassez de habilidades em áreas profissionais cruciais (por exemplo, análise estatística). Deve ser feita dotação orçamentária para prováveis aumentos salariais durante o período da avaliação (dois a três anos, normalmente), para compensar a inflação e para eventos inesperados (contingências).

Algumas avaliações nacionais não alcançaram seus objetivos básicos porque o orçamento foi insuficiente. Embora o orçamento global seja responsabilidade do Ministério da Educação, pessoas com experiência em cálculo de custos e em projetos que envolvem dados em grande escala devem participar das discussões sobre o orçamento. Funcionários

do ministério que não estejam familiarizados com projetos desse porte provavelmente terão dificuldade de compreender a necessidade de orçamento para atividades como provas-piloto e limpeza de dados.

Os números da NAEP dos Estados Unidos fornecem uma ideia geral dos custos que devem ser orçados: coleta de dados (30%), desenvolvimento de instrumentos (15%), análise de dados (15%), apresentação de resultados e divulgação (15%), amostragem (10%), processamento de dados (10%) e gerenciamento (5%) (Ilon, 1996). Em alguns países em que, por exemplo, funcionários do ministério ou integrantes de comissão de exames aplicam a prova como parte de suas funções normais, não se pode fazer uma dotação orçamentária para cobrir algumas atividades. Os custos e salários variarão de acordo com as condições econômicas nacionais. No Camboja (que não está entre os primeiros 100 países do mundo em termos de renda nacional bruta), o pagamento às pessoas que redigiram os itens em 2006 foi equivalente a US\$5 por dia.

Possivelmente, os países com recursos muito limitados não achariam conveniente gastá-los com uma avaliação nacional, especialmente diante da probabilidade de que seus sistemas educacionais tenham muitas necessidades não atendidas. Se realmente quiserem desenvolver uma atividade de avaliação nacional, seria muito aconselhável que limitassem o número de áreas de currículo (apenas uma, em um único nível de ensino) e buscassem a assistência técnica e o apoio de doadores.

Ao considerar os custos, é bom ter em mente que o custo dos programas de prestação de contas e demonstração de resultados – e das avaliações nacionais, em particular – é muito pequeno em comparação com o de outros programas educacionais (veja Hoxby, 2002). O ônus de não realizar uma avaliação – de não descobrir o que está funcionando e o que não está funcionando no sistema educacional – é provavelmente muito maior que o custo de uma avaliação. O Volume 3 discute questões relativas ao custeio de uma avaliação nacional.

Responsabilidade de estimar os componentes de custo de uma avaliação nacional: Ministério da Educação, com insumos de consultoria.

RESUMO DAS DECISÕES

A Tabela 4.5 identifica os órgãos que têm a responsabilidade principal de tomar decisões relativas aos 12 componentes de uma avaliação nacional discutidas neste capítulo.

TABELA 4.5

Órgãos com a Responsabilidade Principal pelas Decisões numa Avaliação Nacional

| Decisão | Responsabilidade principal | | | |
|--|----------------------------|----------------------------------|---------|--------|
| | Ministério da Educação | Comissão de Coordenação Nacional | Agência | Outros |
| Dar orientação política | • | | | |
| Realizar avaliação nacional | | | • | |
| Aplicar testes e questionários | | | • | |
| Escolher a população a ser avaliada | • | • | | |
| Determinar amostra ou população-alvo | • | | | |
| Decidir o que avaliar | • | • | • | |
| Decidir como o aproveitamento é avaliado | | | • | |
| Determinar a frequência da avaliação | • | | | |
| Selecionar métodos de comunicar resultados | | • | • | |
| Determinar procedimentos estatísticos | | | • | |
| Identificar métodos de comunicar e utilizar resultados | • | • | • | • |
| Estimar componentes de custo | • | | | • |

Fonte: Autores.

CAPÍTULO

5

QUESTÕES ENVOLVIDAS NAS FASES DE PROJETO, IMPLEMENTAÇÃO, ANÁLISE, APRESENTAÇÃO DE RELATÓRIO E UTILIZAÇÃO DE UMA AVALIAÇÃO NACIONAL

Neste capítulo, identificamos uma série de questões que determinam a confiança que os interessados possam ter nos resultados de uma avaliação nacional. Para cada um dos cinco componentes da atividade de avaliação nacional (projeto, execução, análise dos dados, elaboração de relatórios, divulgação e utilização dos resultados obtidos), sugerimos uma série de atividades que reforçarão a confiança e que, portanto, devem contribuir para a utilização ótima dos resultados e das constatações. Para cada componente, também identificamos os erros comuns que têm ocorrido em avaliações nacionais e dos quais os avaliadores devem estar conscientes, a fim de evitá-los.

PROJETO

O projeto da avaliação define os parâmetros gerais do exercício: os aproveitamentos a serem avaliados, a série ou a idade em que os alunos serão avaliados, as questões de políticas a resolver, e se a avaliação envolverá toda a população-alvo ou uma amostra de alunos.

Atividades Recomendadas

- Envolver, desde o início, os formuladores de políticas do alto escalão para garantir apoio político e obter sua ajuda no momento de definir o marco de referência da avaliação.
- Determinar e levar em conta as necessidades de informação dos formuladores de políticas durante o processo de seleção dos aspectos do currículo, dos níveis de ensino e dos subgrupos da população-alvo (por exemplo, por região ou por gênero) que serão avaliados.
- Obter apoio dos professores, envolvendo representantes dos professores em decisões de políticas relacionadas com a avaliação.
- Estar ciente de que, se houver fortes interesses em jogo associados ao desempenho dos alunos, isso pode levar à oposição de professores e ao estreitamento do currículo efetivo, na medida em que os professores concentrarão o ensino naquilo que será avaliado.

Erros Comuns

- Deixar de fazer a provisão financeira adequada para os aspectos fundamentais de uma avaliação nacional, incluindo a elaboração de relatórios e sua divulgação.
- Deixar de criar uma comissão de coordenação nacional e de usá-la como fonte de informação e orientação no decurso da avaliação.
- Deixar de obter o comprometimento do governo com o processo de avaliação nacional, o que resulta em: incapacidade de identificar as questões políticas fundamentais que devem ser abordadas na fase de projeto da avaliação; ausência de uma comissão de coordenação nacional; ou realização simultânea de mais de uma avaliação nacional (muitas vezes, apoiadas por doadores externos).
- Deixar de envolver partes interessadas relevantes (por exemplo, os representantes dos docentes ou os formadores de professores) no planejamento da avaliação nacional.
- Omitir da população-alvo avaliada determinado subgrupo (por exemplo, alunos de escolas particulares ou alunos de escolas pe-

quenas), o que poderá introduzir sérias distorções nos resultados da avaliação.

- Definir metas irrealistas para as pontuações do teste (por exemplo, aumento de 25% nas pontuações num período de quatro anos).
- Deixar de alocar tempo suficiente para a elaboração da prova.

IMPLEMENTAÇÃO

A implementação abrange uma vasta gama de atividades, que vão desde o desenvolvimento de instrumentos de avaliação adequados, a seleção dos alunos que responderão aos instrumentos, até a aplicação dos instrumentos nas escolas.

Atividades Recomendadas

- Descrever em detalhes o conteúdo e as habilidades cognitivas que caracterizam o aproveitamento, bem como as variáveis contextuais que serão avaliadas.
- Confiar a elaboração das provas a pessoal familiarizado tanto com os padrões curriculares quanto com os níveis de aprendizado dos alunos (especialmente professores em exercício).
- Utilizar instrumentos que avaliem adequadamente as habilidades e os conhecimentos sobre os quais se requerem informações e que forneçam dados relevantes sobre os subdomínios de habilidades ou conhecimentos (por exemplo, resolução de problemas), em vez de apresentar apenas uma pontuação global.
- Desenvolver itens claros e inequívocos para compor o teste e o questionário, e apresentá-los de forma clara e atraente.
- Assegurar que existam procedimentos adequados para avaliar a equivalência das versões caso haja necessidade de tradução dos instrumentos.
- Fazer um teste-piloto dos itens, questionários e manuais.
- Rever os itens para identificar possíveis ambiguidades e preconceitos relacionados às características dos alunos (por exemplo, gênero, loca-

lização ou pertencimento a grupo étnico), corrigi-los e, se necessário, eliminá-los.

- Fazer uma cuidadosa revisão gráfica de todos os materiais.
- Estabelecer procedimentos para garantir a segurança de todos os materiais da avaliação nacional (por exemplo, provas e questionários) durante todo o processo de avaliação, evitando que caiam em mãos de pessoas não autorizadas.
- Garantir os serviços de uma pessoa ou uma unidade com experiência em amostragem.
- Especificar a população-alvo definida (a população da qual se vai extrair uma amostra – ou seja, a base da amostragem) e a população-alvo excluída (por exemplo, os elementos da população-alvo aos quais é muito difícil chegar ou que não seriam capazes de responder ao instrumento). Devem ser fornecidos dados exatos sobre as populações excluídas.
- Certificar-se de que a amostra proposta é representativa e de tamanho suficiente para fornecer informações sobre as populações-alvo, e com um nível de erro aceitável.
- Selecionar membros da amostra, a partir da base de amostragem, de acordo com as probabilidades conhecidas de seleção.
- Seguir um procedimento padrão quando aplicando provas e questionários. Preparar um manual de aplicação.
- Certificar-se de que os aplicadores de prova estejam completamente familiarizados com o conteúdo das provas, dos questionários e manuais e com os procedimentos de aplicação.
- Elaborar e implementar um mecanismo de garantia de qualidade que cubra, entre outras coisas, validade da prova, amostragem, impressão, aplicação da prova e preparação de dados.

Erros Comuns

- Atribuir as tarefas de elaboração de provas a pessoas não familiarizadas com os níveis prováveis de desempenho dos alunos (por exemplo, acadêmicos), o que pode resultar em provas muito difíceis.

- Espelhar o currículo de forma inadequada nas provas, como, por exemplo, deixando de incluir aspectos importantes do currículo.
- Deixar de incluir certos itens no teste-piloto ou testar os itens numa amostra não representativa da população-alvo.
- Utilizar na versão final da prova um número insuficiente dos itens que compunham o teste-piloto.
- Deixar de dar uma definição clara do conceito avaliado (por exemplo, leitura).
- Incluir um número insuficiente de itens da amostra para alunos que não estão familiarizados com o formato de teste.
- Não incentivar os alunos a solicitar esclarecimentos ao supervisor de teste antes de fazer a prova.
- Deixar de dar instruções adequadas aos impressores de provas, questionários e manuais a respeito de prazos.
- Não fazer uma revisão cuidadosa das provas, dos questionários e manuais de aplicação.
- Usar dados nacionais inadequados ou obsoletos sobre o número de alunos e de escolas para fazer a amostra.
- Deixar de realizar os procedimentos adequados de amostragem, inclusive a seleção de um percentual predeterminado de escolas (por exemplo, 5%).
- Dar treinamento inadequado aos aplicadores da prova e do questionário.
- Permitir intervenção externa (por exemplo, a presença do diretor na sala de aula) durante a aplicação da prova.
- Permitir que os alunos se sentem perto uns dos outros durante a avaliação (isso incentiva a cola).
- Deixar de estabelecer o hábito de se fazerem horas extras, se necessário, para concluir as tarefas-chave a tempo.

ANÁLISE

As análises estatísticas organizam, resumem e interpretam os dados coletados nas escolas. Devem abordar as questões de políticas identificadas na fase de projeto da avaliação nacional.

Atividades Recomendadas

- Garantir serviços estatísticos competentes.
- Preparar um livro de código com instruções específicas para a preparação de dados para análise.
- Conferir e limpar os dados para remover erros (por exemplo, relativos a números, notas fora do intervalo e incompatibilidades entre dados coletados em diferentes níveis).
- Calcular erros de amostragem, levando em conta as complexidades da amostra, tais como estratificação e agrupamentos (*clustering*).
- Pesar os dados para que a participação de cada setor da amostra nas pontuações agregadas de aproveitamento reflita suas proporções na população-alvo.
- Identificar o percentual de alunos que alcançaram os níveis ou padrões previamente definidos como aceitáveis.
- Analisar os dados da avaliação para identificar os fatores que poderiam explicar a variação nos níveis de aproveitamento dos alunos, com vistas a dar subsídios para o processo de formulação de políticas.
- Analisar os resultados por domínios de currículo. Fornecer informações sobre os subdomínios de uma área de currículo (por exemplo, aspectos de leitura, matemática).
- Levantar em conta que inúmeros fatores de medição, curriculares e sociais podem explicar o desempenho dos alunos.

Erros Comuns

- Usar análises estatísticas inadequadas, inclusive deixando de dar peso aos dados da amostra na análise.
- Basear os resultados em números pequenos (por exemplo, uma minoria de professores que integraram a amostra e que poderiam ter respondido a determinada questão).
- Comparar o desempenho dos alunos em diferentes áreas de currículo e afirmar, com base nas diferenças de pontuação média, que os alunos estão se saindo melhor em determinada área.

- Deixar de enfatizar a natureza arbitrária dos pontos de corte das pontuações selecionadas (por exemplo, domínio *versus* não domínio, aprovação *versus* fracasso), dicotomizando resultados e não reconhecendo a vasta gama de pontuações registradas nas provas de um grupo.
- Não reportar erros-padrão associados a estatísticas individuais.
- Computar e divulgar classificações de escolas com base nos resultados de um teste de aproveitamento sem levar em conta fatores-chave contextuais que contribuem para a classificação. Surgem classificações diferentes quando os desempenhos das escolas são comparados usando-se pontuações de desempenho não ajustadas, pontuações ajustadas por fatores contextuais (por exemplo, o percentual de alunos de baixa condição socioeconômica) e pontuações ajustadas por aproveitamento anterior.
- Inferir causalidade onde talvez não possa ser justificada (por exemplo, atribuindo as diferenças no aproveitamento do aprendizado a uma única variável, como a administração de escolas particulares ou o tamanho da turma).
- Comparar os resultados de provas realizadas em dois períodos, embora tenham sido usados itens não equivalentes.
- Comparar os resultados de provas realizadas em dois períodos sem explicitar a medida em que importantes condições contextuais (por exemplo, currículo, matrículas, renda familiar ou o nível de conflitos civis) podem ter mudado nesse ínterim. Embora a maioria das variáveis relacionadas com a educação tenda a não apresentar mudanças muito rápidas durante um curto período (por exemplo, 3 a 4 anos), alguns países adotaram políticas que resultaram em grandes mudanças no número de matrículas. Após a abolição das taxas escolares, por exemplo, o número de alunos matriculados nas escolas aumentou muito no Malawi e em Uganda.
- Limitar a análise principalmente a uma lista de pontuações médias de regiões geográficas ou administrativas.

ELABORAÇÃO DO RELATÓRIO

Faz pouco sentido realizar uma avaliação nacional se os resultados não forem apresentados com clareza e não tiverem relação com as necessidades dos vários interessados.

Atividades Recomendadas

- Elaborar relatórios em tempo hábil tendo em mente as necessidades dos clientes e apresentá-los num formato que seja facilmente compreendido pelas partes interessadas, especialmente aquelas que se encontram em posição de tomar decisões.
- Apresentar um relatório de resultados por gênero e região, se o processo de amostragem assim o permitir.
- Proporcionar informação adequada no relatório ou em um manual técnico para possibilitar a replicação da avaliação.

Erros Comuns

- Escrever relatórios excessivamente técnicos.
- Deixar de destacar alguns resultados principais.
- Fazer recomendações sobre uma variável específica, embora a análise haja questionado a validade dos dados relativos a essa variável.
- Deixar de relacionar os resultados da avaliação a currículo, livros didáticos e aspectos da formação de professores.
- Não reconhecer que fatores fora do controle do professor e da escola contribuem para as pontuações do teste.
- Deixar de reconhecer que as diferenças entre pontuações médias podem não ser estatisticamente significativas.
- Produzir o relatório com atraso, quando já não seja mais possível influenciar decisões relevantes de políticas.
- Fazer uma revisão da literatura excessivamente extensa no relatório de avaliação.

- Deixar de divulgar junto ao público interessado um texto em separado com as mensagens mais relevantes do relatório.

DIVULGAÇÃO E UTILIZAÇÃO DAS CONSTATAÇÕES

É importante que os resultados das avaliações nacionais não fiquem esquecidos nas prateleiras dos formuladores de políticas, mas que sejam transmitidos numa linguagem adequada a todos os que podem afetar a qualidade do aprendizado dos alunos.

Atividades Recomendadas

- Disponibilizar os resultados aos interessados, especialmente aos principais formuladores e gestores de políticas.
- Usar os resultados, quando apropriado, para a elaboração de políticas e para melhorar o ensino e os currículos.

Erros Comuns

- Ignorar os resultados durante o processo de formulação de políticas.
- Deixar de considerar as implicações dos resultados da avaliação nacional para as partes interessadas relevantes (por exemplo, os formadores de professores ou o pessoal que trabalha com currículos).
- Deixar de refletir com a equipe da avaliação nacional sobre as lições aprendidas e não levá-las em conta durante as avaliações posteriores.

CAPÍTULO

6

AVALIAÇÕES INTERNACIONAIS DO APROVEITAMENTO DE ALUNOS

Neste capítulo, descrevemos as avaliações internacionais do aproveitamento educacional dos alunos, considerando-se que são usadas em muitos países como fontes de dados para uma avaliação nacional. Primeiro, delineamos as principais características das avaliações internacionais em termos de suas semelhanças e diferenças com as avaliações nacionais. Em seguida, descrevemos o aumento da atividade de avaliação internacional. Então, identificamos as vantagens das avaliações internacionais, bem como os problemas associados a elas.

Uma avaliação internacional do aproveitamento dos alunos tem muitas semelhanças com uma avaliação nacional. Ambos os exercícios adotam procedimentos similares (nas etapas de construção de instrumentos, amostragem, pontuação e análise). Também podem ter objetivos semelhantes, como: determinar quão bem os alunos estão aprendendo no sistema educacional; identificar pontos fortes e fracos específicos nos conhecimentos e nas habilidades adquiridos pelos alunos; comparar os aproveitamentos de subgrupos da população-alvo (por exemplo, definidos em termos de gênero ou lugar); ou determinar a relação entre o aproveitamento dos alunos e uma variedade de características do ambiente de aprendizado da escola, das famílias e das comunidades. Além

disso, ambos os exercícios podem tentar estabelecer se o aproveitamento dos alunos muda ao longo do tempo (Kellaghan e Greaney, 2004). Na prática, porém, nem sempre é clara a razão que leva um país a decidir participar de uma avaliação internacional (Ferrer, 2006).

A principal vantagem de uma avaliação internacional em relação a uma avaliação nacional é que a primeira tem como objetivo fornecer aos formuladores de políticas, educadores e ao público em geral informações sobre o sistema educacional de seus respectivos países em comparação com os de outros (Beaton e outros, 1999; Husén 1973; Postlethwaite, 2004). Presume-se que essa informação pressione os formuladores de políticas e os políticos para que melhorem os serviços. Além disso, espera-se que a informação contribua para a maior compreensão dos fatores (que variam de país para país) relacionados com as diferenças de aproveitamento dos alunos.

As áreas de currículo que têm atraído a maior taxa de participação em estudos internacionais ao longo dos anos são compreensão em leitura, matemática e ciências. Têm sido realizados estudos em escolas de primeiro e segundo grau. Normalmente, usa-se uma combinação de série escolar e idade para determinar quem irá participar (por exemplo, os alunos de duas séries adjacentes que contenham as maiores proporções nas idades de 9 a 13 anos; alunos dos níveis de ensino que contenham o maior número nas idades de 9 e 14 anos; a série mais elevada de duas séries adjacentes que contenham o maior número de alunos com 9 anos). Em outro estudo internacional, foram selecionados alunos de determinada idade (15 anos).

Os resultados de avaliações internacionais como o TIMMS (Tendências no Estudo Internacional de Matemática e Ciências) e o PISA (Programa Internacional de Avaliação de Alunos), e também os de avaliações regionais, têm sido usados para preparar relatórios nacionais específicos sobre o desempenho de cada país. Para realizar essas análises, as bases de dados internacionais podem ser acessadas.

Existem variações consideráveis entre os países no que se refere à extensão do uso dos resultados de avaliações internacionais e nacionais para a formulação de políticas. Muitos países industriais conduzem as próprias avaliações nacionais e também participam de avaliações in-

ternacionais. Os Estados Unidos têm a própria Avaliação Nacional do Progresso Educacional para as 4^a, 8^a e 12^a séries; também participam de avaliações internacionais de aproveitamento. Alguns países industrializados têm participado de avaliações internacionais, mas não realizam avaliações nacionais (por exemplo, a Federação Russa e a Alemanha). Da mesma forma, alguns países em desenvolvimento têm usado as avaliações internacionais como fonte única de suas avaliações nacionais (Braun e Kanjee, 2007). Muitos dos países mais pobres do mundo não têm tomado parte em avaliações internacionais nem realizado avaliações nacionais, embora a situação esteja mudando nos últimos anos.

CRESCIMENTO DA ATIVIDADE DE AVALIAÇÃO INTERNACIONAL

A atividade de avaliação internacional teve início em 1958, quando um grupo de pesquisadores reuniu-se para examinar a possibilidade de se realizar um estudo de resultados de aproveitamento (que já haviam sido medidos) e de seus determinantes, comparando escolas dentro de um mesmo país e entre países (Husén e Postlethwaite, 1996). Desde então, mais de 60 países têm participado de estudos internacionais de aproveitamento em uma área de currículo ou em várias: leitura, matemática, ciências, escrita, literatura, línguas estrangeiras, educação cívica e alfabetização digital. As avaliações internacionais mais conhecidas são o TIMSS (veja B.1 no Apêndice B), o PIRLS, realizado pela IEA (veja B.2 no Apêndice B) e o PISA, realizado pela OCDE (veja B.3 no Apêndice B). Avaliações regionais em leitura e matemática têm sido feitas em países da África Austral e da África Oriental (veja C.1 do Apêndice C), na África francófona (veja C.2 do Apêndice C) e na América Latina (veja C.3 no Apêndice C). Diversas características nas quais diferem o TIMSS e o PISA são apresentadas na Tabela 6.1 (veja também B.1 e B.3 no Apêndice B).

O número de países que participam de estudos internacionais tem aumentado ao longo dos anos. Até a década de 1980, usualmente os estudos abrangiam menos de 20 países, mas o Estudo sobre Letramento em Leitura realizado pela IEA em 1991 atraiu 32 países. Em 2003, 52

TABELA 6.1
Comparação entre o TIMSS e o PISA

| | TIMSS 2003 | PISA 2003 |
|------------------------|--|---|
| Objetivos | <p>Fornecer evidências comparativas sobre em que medida os alunos dominam o conteúdo do currículo oficial da escola em matemática e ciências, com um a uma série de países.</p> <p>Monitorar mudanças nos níveis de aproveitamento ao longo do tempo.</p> <p>Monitorar as atitudes dos alunos com relação a matemática e ciências.</p> <p>Analisar a relação entre uma série de fatores instrucionais e escolares e níveis de aproveitamento.</p> <p>(A leitura é coberta numa avaliação à parte, feita pelo PIRLS.)</p> | <p>Fornecer evidências comparativas sobre o "rendimento" do sistema de ensino nos principais países industrializados e avaliar se os alunos podem aplicar seus conhecimentos e competências em leitura, matemática e ciências a situações da vida real.</p> <p>Monitorar mudanças nos níveis de aproveitamento e na equidade dos resultados do aprendizado ao longo do tempo.</p> <p>Monitorar as abordagens do aluno à aprendizagem e suas atitudes em relação a matemática, ciências e leitura.</p> <p>Fornecer um banco de dados para a elaboração de políticas.</p> |
| Marco de referência | Desenvolvido por especialistas em conteúdo de alguns países participantes. | Desenvolvido por especialistas em conteúdo de alguns países participantes. |
| População-alvo | 4 ^a e 8 ^a séries | Alunos com 15 anos |
| Adequação do currículo | Projetado para avaliar o currículo oficial, organizado em torno de áreas de currículo reconhecidas e comuns aos países participantes. | Projetado para cobrir os conhecimentos adquiridos na escola e fora dela, definidos em termos de ideias e competências abrangentes aplicadas a situações pessoais, educacionais, profissionais, públicas e científicas. |

| | | | |
|--|--|--|--|
| Diferenças no conteúdo de itens (matemática, 8ª série) | 8ª série, distribuição de itens: <ul style="list-style-type: none"> • Número, 30% • Álgebra, 25% • Dados, 15% • Geometria, 15% • Mensuração, 15% | Matemática, ideias globais: <ul style="list-style-type: none"> • Quantidade • Espaço e forma • Mudança e relações • Incerteza | Distribuição dos itens: <ul style="list-style-type: none"> • Número, 31,8% • Geometria, 21,2% • Estatística, 21,2% • Funções, 10,6% • Matemática discreta, 5,9% • Probabilidade, 5,9% • Álgebra, 3,5% |
| Processos cognitivos | 8ª série: <ul style="list-style-type: none"> • Solução de problemas de rotina: 40% • Uso de conceitos: 20% • Conhecimento de fatos e procedimentos: 5% • Raciocínio: 25% | Distribuição do item: <ul style="list-style-type: none"> • Associação: 47% • Reprodução: 31% • Reflexão: 22% | |
| Tipos de item (matemática) | Cerca de dois terços são itens de múltipla escolha, e o restante são questões fechadas ou abertas. | Cerca de um terço são itens de múltipla escolha e o restante, em geral, são questões fechadas (uma possível resposta certa) ou abertas (mais de uma possível resposta certa). | |
| Frequência | A cada quatro anos: a mesma ênfase em matemática e ciências em cada ciclo. | A cada três anos: ampla cobertura de um domínio (assunto) a cada nove anos (leitura em 2000, matemática em 2003 e ciências em 2006), além de cobertura menos extensa dos outros dois a cada três anos. | |
| Cobertura geográfica | 48 países: 20 de renda alta, 26 de renda média e dois de renda baixa. | 30 países da OCDE e outros 11 países. | |
| Análise | Quatro níveis de referência e uma pontuação média, com base nos resultados de todos os países participantes. | Sete níveis de proficiência em matemática e uma pontuação média, baseados em países da OCDE. | |

Fonte: Marcos de referência do TIMSS e do PISA; Centro Nacional de Estatísticas Educacionais dos Estados Unidos, s.d.; base de dados Indicadores de Desenvolvimento Mundial.

países participaram do TIMSS e 41 do PISA (30 Estados-membros da OCDE e 11 países “parceiros”). Além disso, os estudos internacionais dos últimos anos têm destacado o acompanhamento do desempenho ao longo do tempo. As três mais importantes avaliações internacionais hoje existentes (TIMSS, PIRLS e PISA) são aplicadas a intervalos regulares e passaram a ser descritas como estudos de tendências.

Como regra, a participação dos países não industrializados nos estudos internacionais tem sido baixa. No entanto, esse número também vem aumentando ao longo dos anos. O TIMSS atraiu o maior número deles em 2003 (sete países da África) e em 2007 (seis da África). Como ocorreu, de modo geral, nos estudos internacionais, os países não industrializados também mostraram maior interesse em participar de avaliações de matemática e leitura que de outras áreas de currículo.

O crescimento recente da participação em estudos internacionais pode ser atribuído à globalização, a um movimento nas áreas de saúde e educação para comparar e avaliar os serviços tomando como padrão de referência aqueles de outros países e o interesse global. Algumas evidências de pesquisa apoiam a ideia de que a qualidade da educação (em particular, naqueles aspectos representados pelo aproveitamento em matemática e ciências) desempenha papel importante no crescimento econômico, embora essa relação não seja inteiramente consistente entre os países ou ao longo do tempo (Coulombe, Tremblay e Marchand, 2004; Hanushek e Kimko, 2000; Hanushek e Wössmann, 2007; Ramirez *et al.*, 2006). Seja qual for a razão, as políticas de educação em todo o mundo têm se centrado, cada vez mais, na necessidade de acompanhar o aproveitamento agregado dos alunos num contexto internacional.

VANTAGENS DAS AVALIAÇÕES INTERNACIONAIS

Diversas razões têm sido propostas aos países para incentivá-los a participar de uma avaliação internacional do aproveitamento dos alunos. Talvez a mais óbvia seja que os estudos internacionais fornecem um marco de referência a partir do qual é possível avaliar o aproveitamento dos alunos e os currículos de um país e implantar procedimentos para

lidar com as deficiências percebidas (Strauss, 2005). Ao comparar os resultados de diferentes países, um país pode utilizar a avaliação para ajudá-lo a definir o que é possível alcançar, saber como o aproveitamento está distribuído e ver as relações entre o aproveitamento médio e sua distribuição. Por exemplo, será que um elevado aproveitamento médio pode coexistir com baixas disparidades no desempenho? Os resultados do PISA sugerem que sim.

Os dados sobre aproveitamento fornecem apenas informações limitadas. Tem-se argumentado a favor dos estudos internacionais que eles podem se beneficiar com a variabilidade existente entre os sistemas educacionais: a diversidade de condições que podem ser estudadas é mais ampla que a encontrada em um único país (Husén, 1973). Com base nisso, a análise dos dados coletados nesses estudos normalmente considera as associações existentes entre o aproveitamento e uma vasta gama de variáveis contextuais. As variáveis consideradas incluem conteúdo do currículo, tempo dedicado a trabalhos escolares, treinamento de professores, tamanho da turma e organização do sistema educacional. Claramente, o valor dos estudos internacionais é reforçado, na medida em que permitem a pesquisadores e formuladores de políticas elaborar hipóteses sobre o que explicaria as diferenças entre países quanto ao aproveitamento dos alunos. Além disso, os estudos internacionais também fornecem uma base para a avaliação de políticas e práticas.

As avaliações internacionais têm o potencial de trazer à luz alguns conceitos que, embora necessários para se compreender a educação, estão sendo negligenciados em um país – por exemplo, a definição de letramento, ou a conceituação do currículo em termos de intenção, implementação e aproveitamento (ver, por exemplo, Elley, 2005). As avaliações também podem ajudar a identificar pressupostos que talvez tenham sido adotados sem maiores exames (por exemplo, o valor da educação abrangente, em comparação com a educação seletiva; turmas menores sendo associadas a maior aproveitamento; ou a repetência vista como um benefício para os alunos).

Os estudos internacionais podem atrair a atenção da mídia e de um amplo espectro de partes interessadas, tais como políticos, formuladores de políticas, acadêmicos, professores e o público em geral. As diferenças

nos níveis de aproveitamento entre os países são evidentes nas estatísticas descritivas fornecidas nos relatórios desses estudos. De fato, essas diferenças são geralmente destacadas nas tabelas classificativas em que os países aparecem ordenados pelo nível médio de aproveitamento. Os dados comparativos fornecidos por esses estudos causam mais impacto (e mais “choque”) que os resultados de uma avaliação nacional. Os resultados insatisfatórios podem incentivar o debate, e esse, por sua vez, pode fornecer aos políticos e a outros formuladores de políticas uma justificativa para aumentar os recursos orçamentários destinados à educação – especialmente se os resultados insatisfatórios estiverem associados a um nível mais baixo de gastos na área.

Uma característica importante de uma avaliação internacional é o fato de fornecer dados que podem ser usados por cada país para fazer análises internas e produzir o próprio relatório de avaliação nacional. Essa prática é seguida pelos países que participam do PISA (veja B.3 no Apêndice B) e do SACMEQ (veja C.1 no Apêndice C). A prática é reforçada se, indo além dos dados coletados para o estudo internacional, também forem obtidos dados sobre questões que constituam uma preocupação ou um interesse relevante de países específicos.

A participação em avaliações internacionais tem uma série de vantagens práticas, especialmente para os países nos quais as universidades não conseguem desenvolver os tipos de habilidades necessárias para se realizar uma avaliação nacional. Em primeiro lugar, uma agência central pode realizar análises em nível nacional a serem utilizadas nos relatórios de cada país. Em segundo lugar, os estudos podem contribuir para o desenvolvimento da capacidade local em diversas atividades técnicas: amostragem, definição de aproveitamento, elaboração de provas, análise de estatísticas e elaboração de relatórios. Em terceiro lugar, as necessidades de pessoal e os custos envolvidos (por exemplo, para o desenvolvimento de instrumentos, limpeza e análise de dados) podem ser inferiores aos de avaliações nacionais porque certos custos são compartilhados com outros países.

Um estudo dos efeitos do TIMSS sobre o ensino e o aprendizado de matemática e ciências nos países participantes fornece evidências da diversidade de atividades que um estudo internacional pode gerar (Rorbitaille, Beaton e Plomp, 2000):

- Os resultados do TIMSS foram destacados em debates parlamentares sobre mudanças planejadas na política de educação (Japão).
- O ministro da Educação estabeleceu uma força-tarefa em matemática e ciências (Nova Zelândia).
- O presidente ordenou que fosse implementado um “pacote de resgate” para melhorar o desempenho em ciências e matemática (no qual o treinamento de professores receberia uma atenção especial) (Filipinas).
- Foram estabelecidos padrões de referência nacionais em letramento em leitura e letramento matemático (Austrália).
- Os resultados contribuíram para o desenvolvimento de novos padrões educacionais em matemática e ciências (Federação Russa).
- Os resultados ajudaram a mudar a natureza das discussões públicas na área de educação, deixando de lado as discussões baseadas em opiniões e privilegiando-se discussões baseadas em fatos (Suíça).
- Os resultados levaram ao desenvolvimento de materiais de instrução derivados da análise dos equívocos e erros mais comuns cometidos pelos alunos ao responder às questões do TIMSS (Canadá).
- Os resultados aceleraram mudanças na revisão de currículos (República Tcheca; Cingapura).
- Os resultados do TIMSS foram identificados como um dos diversos fatores que influenciam mudanças na política de educação em matemática (Inglaterra).
- Foram formadas comissões para rever currículos de matemática e ciências (Kuwait).
- Novos tópicos foram adicionados ao currículo de matemática (Romênia).
- Introduziu-se novo conteúdo nos currículos de matemática e ciências relacionados com situações da vida real (Espanha).
- Os resultados ajudaram a destacar a necessidade de melhorar o equilíbrio entre matemática pura e matemática aplicada a contextos (Suécia).
- Os resultados do TIMSS destacaram crenças sobre diferenças de gênero e atitudes negativas com relação à ciência e à matemática e foram utilizados como base para uma reforma curricular e para o desenvolvimento profissional dos professores (Coreia do Sul).

- Os resultados influenciaram o produto das discussões sobre como melhorar a organização da formação de professores e como dar maior ênfase à questão (Islândia).
- Os resultados do TIMSS levaram a medidas para reforçar o desenvolvimento profissional de professores de matemática e ciências (Noruega, Estados Unidos).
- Criou-se um sistema centralizado de exames, em parte como resposta aos resultados do TIMSS (Letônia).
- Os resultados do TIMSS influenciaram grandes mudanças no ensino, na organização das escolas e das turmas, na formação de professores e na definição de metas para as escolas (Escócia).
- Os resultados do TIMSS afetaram a pesquisa educacional, o desenvolvimento de padrões, o desenvolvimento de documentos curriculares, os estudos dos professores, as metodologias de ensino de matemática e ciências e o desenvolvimento de livros didáticos (Rep. Eslovaca).

Os resultados das análises dos dados do PISA levaram ao seguinte:

- Lançaram dúvidas sobre o valor do uso extensivo de computadores na sala de aula como forma de melhorar o aproveitamento.
- Destacaram o fato de que o nível de gastos nacionais com educação não está associado ao aproveitamento (entre os países participantes).
- Deflagraram amplo debate político sobre a educação (Alemanha).
- Contribuíram para o desenvolvimento do currículo de ciências no ensino fundamental (Irlanda).
- Enfatizaram a complexidade da relação entre situação socioeconômica e o aproveitamento em leitura na comparação entre países.
- Destacaram a ligação entre aproveitamento, tipos de escola e acompanhamento do currículo dentro de escolas.
- Embasaram a noção de que escolas públicas e privadas tendem a ter os mesmos efeitos para os mesmos tipos de alunos, mas que escolas privadas que dependem do governo são relativamente mais efetivas para os alunos de níveis socioeconômicos mais baixos.
- Salientaram a necessidade de programas intensivos de idioma e leitura para alunos estrangeiros, a fim de incrementar o aproveitamento (Suíça).

PROBLEMAS DAS AVALIAÇÕES INTERNACIONAIS

Apesar das vantagens óbvias, vários problemas associados às avaliações internacionais devem ser considerados antes que os países decidam participar de uma (veja Kellaghan, 1996).

O primeiro problema refere-se à dificuldade de se desenhar um procedimento de avaliação que meça adequadamente os resultados de diversos currículos. Embora os currículos em todo o mundo tenham elementos comuns, particularmente no ensino básico, também existem diferenças consideráveis entre os países no que se refere ao que é ensinado, a quando é ensinado e aos padrões de aproveitamento esperados.

A revisão de itens do TIMSS feita pela África do Sul mostra que apenas 18% dos itens de ciência correspondiam ao currículo nacional da 7ª série, enquanto 50% correspondiam ao currículo da 8ª série (Howie e Hughes, 2000). Quanto mais diferirem os países em termos de currículos e níveis de aproveitamento, mais difícil se torna criar um processo de avaliação que sirva a todos os que participam de uma avaliação internacional, e mais duvidosa a validade de quaisquer inferências feitas sobre aproveitamentos comparativos.

Seria de se esperar que um teste de desempenho baseado no conteúdo do currículo nacional fornecesse uma medida mais válida do domínio do currículo que outro concebido para servir como denominador comum dos currículos oferecidos em 30 a 40 países. Por exemplo, uma autoridade nacional encarregada do currículo e os que fazem o projeto de uma avaliação internacional poderiam atribuir pesos bastante diferentes à habilidade de fazer inferências a partir de um texto. Uma avaliação nacional, ao contrário de uma avaliação internacional, também pode testar os aspectos curriculares exclusivos de cada país.

A elaboração de um instrumento de avaliação comum é mais difícil em algumas áreas de currículo (por exemplo, ciências e estudos sociais) que em outras (por exemplo, leitura). No caso de ciências, descobriu-se que os padrões de aproveitamento eram mais heterogêneos que em matemática. Além disso, é necessário recorrer a um maior número de fatores para explicar as diferenças de desempenho dos alunos em ciências do

que em matemática. Assim, é difícil imaginar um teste de ciências que seja adequado para diferentes sistemas educacionais.

Um segundo problema com os estudos internacionais é que, embora estudos anteriores tivessem o objetivo ambicioso de aproveitar a variação existente entre sistemas educacionais para avaliar a importância relativa de uma variedade de recursos escolares e processos de instrução, essa meta, na prática, provou-se muito difícil de alcançar. Dado que o efeito relativo das variáveis depende do contexto em que estão inseridas, não se pode presumir que práticas associadas a alto aproveitamento em um país mostrarão relação semelhante em outro. Na verdade, constatou-se que variam de país para país a força das correlações entre os fatores contextuais ou antecedentes e o aproveitamento dos alunos (ver, por exemplo, OCDE e Instituto de Estatísticas da UNESCO 2003; Wilkins, Zembylas e Travers, 2002). Existem dificuldades específicas quando países em desenvolvimento estão envolvidos num estudo desenhado para países industrializados, pois os fatores socioeconômicos dos primeiros podem ser muito diferentes daqueles que prevalecem nos países industrializados e podem incluir pobreza, aspectos nutricionais e de saúde, infraestrutura educacional precária e carência de financiamento.

O terceiro problema associado às avaliações internacionais refere-se ao fato de que os públicos-alvo e as amostras de alunos que participam das avaliações internacionais podem não ser exatamente comparáveis. Por exemplo, as diferenças de desempenho podem resultar do fato de que difere entre os países a dimensão de certas categorias de alunos que são transferidos para turmas especiais e podem acabar sendo excluídos de uma avaliação (por exemplo, os alunos em programas especiais ou alunos de escolas em que a língua de ensino é diferente da língua da avaliação).

O problema é mais evidente quando a idade para entrar na escola, a taxa de retenção e a taxa de evasão diferem entre os países, e isso é particularmente relevante em estudos dos quais participam países industrializados e países em desenvolvimento. Em alguns desses últimos, grandes proporções de alunos abandonam os estudos bem antes do final do período de escolaridade obrigatória. Enquanto as taxas de es-

colarização líquida no ensino básico na Europa Ocidental e na América do Norte alcançam quase 100%, nos países da África Subsaariana são inferiores a 60%, em média (UNESCO, 2002). Os padrões de abandono precoce da escola podem diferir de país para país. Na América Latina e nos países árabes, os meninos são mais propensos que as meninas a não completar a 5ª série; o inverso é verdadeiro em alguns países africanos (por exemplo, Guiné e Moçambique). Surgiram problemas de amostragem para o TIMSS na República do Iêmen, onde diversas escolas não têm a 4ª série e onde uma escola para crianças nômades não pôde ser localizada.

Numa avaliação nacional também podem surgir problemas de comparabilidade semelhantes. Por exemplo, o diferencial de desempenho dos alunos em vários estados da Índia tem sido atribuído ao diferencial de taxas de sobrevivência (veja A.1 no Apêndice A).

O quarto problema leva em conta o fato de que, como a variação das pontuações do teste é um fator importante quando se pretende descrever adequadamente o aproveitamento dos alunos no sistema escolar e, adicionalmente, determinar os correlatos de aproveitamento, são necessários exames nacionais cuidadosamente concebidos para assegurar uma distribuição relativamente ampla das pontuações. No entanto, muitos itens que aparecem em avaliações internacionais têm representado dificuldades muito grandes para os alunos de países menos industrializados, o que resulta no achatamento (pouca variação) das pontuações dos testes. Isso fica evidente nos dados apresentados na Tabela 6.2, baseados numa seleção entre países que participaram do TIMSS 2003.

Os dados mostram o percentual de alunos da 8ª série que atingiram determinados níveis ou padrões de referência de desempenho quando comparados a todos os alunos que fizeram a prova. Daqueles que fizeram o teste de matemática, 7% alcançaram o padrão de referência internacional “avançado”, 23% alcançaram o padrão “alto”, a metade ficou no padrão “intermediário” e cerca de três quartos ficaram no nível “baixo”. Em forte contraste, nenhum aluno de Gana alcançou os padrões internacionais “avançado” e “alto”, sendo que 2% atingiram o padrão “intermediário” e 9% ficaram no “baixo”.

TABELA 6.2

Porcentagem de Alunos da 8ª Série que Alcançaram os Padrões de Referência Internacionais do TIMSS em Matemática: Países com Pontuações Altas e Baixas

| Países | Avançado ^a | Alta ^a | Intermediária ^a | Baixa ^a |
|----------------------------|-----------------------|-------------------|----------------------------|--------------------|
| Cingapura | 44 | 77 | 93 | 99 |
| Taipei | 38 | 66 | 85 | 96 |
| Coreia do Sul | 35 | 70 | 90 | 98 |
| Média internacional | 7 | 23 | 49 | 74 |
| Filipinas | 0 | 3 | 14 | 39 |
| Bahrain | 0 | 2 | 17 | 51 |
| África do Sul | 0 | 2 | 6 | 10 |
| Tunísia | 0 | 1 | 15 | 55 |
| Marrocos | 0 | 1 | 10 | 42 |
| Botsuana | 0 | 1 | 7 | 32 |
| Arábia Saudita | 0 | 0 | 3 | 19 |
| Gana | 0 | 0 | 2 | 9 |

Fonte: Mullis et al., 2004, 64.

^a Definições usadas no TIMSS, 2003: Avançado: Os alunos podem organizar informações, fazer generalizações, solucionar problemas não rotineiros, extrair conclusões dos dados e justificá-las. Alto: Os alunos podem aplicar sua compreensão e seu entendimento numa grande variedade de situações relativamente complexas. Intermediário: Os alunos podem aplicar conhecimento matemático básico em soluções simples. Baixo: Os alunos têm algum conhecimento matemático básico.

Da mesma forma, no PISA 2003, o uso limitado da avaliação para a formulação de políticas internas ficou sublinhado quando se constatou a falta de variância estatística nas notas dos testes de diversos países participantes; a maioria dos alunos com 15 anos no Brasil, na Indonésia e na Tunísia teve pontuação inferior ao Nível 1. (O Nível 2 tem sido sugerido como um requisito mínimo para que os alunos entrem no mundo do trabalho ou possam continuar os estudos.) Claramente, é muito limitada a informação que os estudos fornecem aos formuladores de políticas e tomadores de decisão sobre o escopo do aproveitamento dos alunos nesses sistemas educacionais. Além disso, devido à pouca variância no aproveitamento, as correlações entre o aproveitamento e as variáveis contextuais ou associadas à escola pouco revelariam sobre os fatores que contribuem para o aproveitamento.

Em quinto lugar, as avaliações internacionais enfrentam um problema quando o foco principal da apresentação dos resultados da avaliação fica concentrado no ordenamento dos países em termos das pontuações mé-

dias dos alunos (tópico que, em geral, atrai o maior interesse da mídia). As classificações em si não nos dizem nada sobre os inúmeros fatores que podem estar na base das diferenças de desempenho entre os países. Além disso, as classificações podem ser enganosas quando se ignora a significância estatística das diferenças médias de aproveitamento. Dependendo dos demais países que integram a lista, a classificação de determinado país pode variar – e esta é uma consideração importante quando se comparam classificações ao longo do tempo. Assim, por exemplo, se em determinado ano diminui o número de países que tradicionalmente têm alta pontuação e aumenta o número dos que costumam apresentar baixos resultados, a classificação de um país pode aumentar sem que isso indique necessariamente melhoria em seu aproveitamento.

O sexto problema refere-se ao fato de que o desempenho insatisfatório numa avaliação internacional (assim como numa avaliação nacional) pode acarretar alguns riscos políticos para funcionários graduados associados à qualidade da educação, inclusive ministros e secretários da Educação. O risco tende a ser maior quando a classificação internacional de um país é menor que a de um tradicional país rival. Em alguns países nos quais os dados foram coletados, os funcionários se recusaram a permitir a inclusão dos resultados em publicações que comparavam países. (A IEA já não permite que os países participantes optem por ficar fora das comparações.) A obtenção de dados comparativos de países vizinhos ou de outros dentro de uma região pareceria preferível a se obterem dados sobre países de todo o mundo, entre os quais existem grandes disparidades de níveis de desenvolvimento socioeconômico. Um exemplo dessa abordagem é encontrado na América Latina e no Caribe, onde 13 países realizaram em conjunto uma avaliação de competências básicas em linguagem e matemática em 1997 (veja C.3 no Apêndice C). As avaliações do SACMEQ realizadas no sul e no leste da África sob os auspícios de uma rede de ministérios na década de 1990 também permitiram comparações internacionais no âmbito regional (veja C.1 no Apêndice C).

Em sétimo lugar na lista de problemas das avaliações internacionais, pode ser muito difícil atender às exigências de cumprimento dos prazos em países que carecem de pessoal administrativo e têm de lidar com precária infraestrutura de comunicações (veja Quadro 6.1).

QUADRO 6.1**A Experiência da África do Sul com Avaliações Internacionais**

A experiência da África do Sul com o TIMSS sublinha os problemas que enfrentam os realizadores de avaliações internacionais. Pode ser difícil, ou mesmo impossível, atender aos prazos impostos pelos organizadores naquelas situações em que os serviços de correios e de telefonia sejam precários ou os fundos de viagem para chegar às escolas sejam insuficientes.

Outros problemas incluem falta de dados exatos sobre o público-alvo e sobre as escolas; baixa capacidade de gestão; insuficiente atenção aos detalhes, especialmente na edição, codificação e coleta de dados; falta de financiamento para apoiar os que trabalham no projeto e dificuldade em garantir um material impresso de qualidade em tempo hábil. Certas instruções para os aplicadores de teste (por exemplo, para que caminhem entre as filas de carteiras durante o exame) são obviamente inadequadas quando as salas de aula não têm corredores.

Fonte: Howie, 2000.

O tempo alocado para a realização de várias tarefas associadas à organização de uma avaliação internacional (por exemplo, a impressão ou distribuição dos cadernos de prova), que pode ser considerado razoável nos países industrializados, talvez seja insuficiente em muitos países em desenvolvimento que enfrentam inúmeros problemas básicos, inclusive a precariedade dos sistemas de comunicação.

Finalmente, existem custos substanciais associados à participação em um estudo internacional. Um país que desejasse participar do TIMSS para a 8ª série deveria pagar US\$40 mil, além de todos os custos relativos a impressão, distribuição, aplicação da prova, entrada de dados e pontuação. As avaliações nacionais, é claro, também envolvem custos consideráveis.

CONCLUSÃO

Os leitores que perseveraram até este ponto já deverão estar familiarizados com as principais características das avaliações nacionais e internacionais, as semelhanças entre as avaliações e os aspectos nos quais diferem, as razões para se engajar em uma avaliação e os problemas aos quais prestar atenção ao longo do processo. Já deverão ter também uma compreensão geral das principais atividades envolvidas, incluindo identificação das questões-chave de políticas ligadas a educação, construção de instrumentos, seleção das escolas e dos alunos que irão representar o sistema educacional, análise de dados para descrever o aproveitamento dos alunos e seus correlatos, e comunicação de resultados a diferentes públicos. Todas essas tarefas requerem conhecimento especializado e habilidades específicas.

Se o leitor é um formulador ou gestor de políticas num Ministério da Educação, é pouco provável que possua qualquer dos conhecimentos ou habilidades especializados envolvidos nos detalhes da execução de uma avaliação nacional. Mas isso não significa que não tenha papel crucial a desempenhar numa avaliação em suas diversas etapas – deflagrar o processo e sua concepção geral, atuar como facilitador durante a implementação, a interpretação e a aplicação das conclusões. Neste capítulo,

prestamos atenção especial ao papel do formulador ou gestor de políticas nas etapas de desenvolvimento e institucionalização da atividade de avaliação nacional e destacamos sua participação no esforço de utilizar o mais plenamente possível os resultados.

Formuladores ou gestores de políticas aos quais cabe decidir se será realizada uma avaliação nacional (ou se o país participará de uma avaliação internacional) devem estar persuadidos de que as informações resultantes da avaliação serão úteis para identificar problemas no sistema educacional e embasar as políticas e práticas destinadas a solucioná-los. O compromisso desses protagonistas provavelmente será reforçado se a avaliação reunir cinco condições.

Em primeiro lugar, os aproveitamentos avaliados devem ser considerados resultados importantes da escolaridade e refletir adequadamente o currículo. Em segundo, o instrumento utilizado na avaliação deve ter o potencial de fornecer informações diagnósticas sobre aspectos do aproveitamento dos alunos, particularmente sobre os pontos fortes e fracos no perfil de aproveitamento. Em terceiro lugar, o método de amostragem (quando se tratar de uma avaliação baseada numa amostra) deve assegurar que os dados coletados representam adequadamente o aproveitamento do sistema educacional como um todo (ou de uma parte dele claramente identificada). Em quarto lugar, devem ser usadas análises apropriadas para identificar e descrever as principais características dos dados, incluindo as relações entre variáveis significativas. Em quinto lugar, os aspectos técnicos da avaliação devem observar os padrões profissionais vigentes em áreas como elaboração de provas, amostragem e análise estatística.

Todas essas atividades exigem consideráveis recursos e apoio político. Por exemplo, o formulador ou gestor de políticas tem o papel crucial de garantir que estejam disponíveis os conhecimentos e habilidades necessários para projetar, gerenciar e interpretar uma avaliação nacional. Em muitos países, eles não existirão, e deverão ser especificamente desenvolvidos para isso. Esse desenvolvimento exigirá programas iniciais de treinamento de longa ou curta duração. Na sequência desses programas, devem ser tomadas providências para aumentar, numa base regular, as habilidades técnicas das pessoas envolvidas na realização de uma avalia-

ção nacional, por meio de programas de formação no país, participação em reuniões profissionais e estudos de graduação de maior duração.

Em alguns países, a atividade de avaliação nacional parece operar à margem do sistema educacional, divorciada da estrutura e dos processos normais de elaboração de políticas e tomada de decisão. Nessa situação, não há garantia de que as informações obtidas numa avaliação serão usadas para orientar as políticas ou que avaliações nacionais serão realizadas no futuro para monitorar mudanças no aproveitamento ao longo do tempo. Para que essas questões sejam tratadas, a atividade de avaliação nacional deve tornar-se parte normal do funcionamento do sistema educacional. Isso exigirá que alguns formuladores de políticas do alto escalão participem ativamente do desenho global da avaliação e estejam presentes (ou representados) na comissão de coordenação nacional. Também requererá orçamento adequado e decisão sobre a localização física e institucional da atividade, que variará de país para país, dependendo das circunstâncias locais.

É muito importante que o governo assuma compromisso de longo prazo com a construção de uma base institucional forte para a realização regular de avaliações nacionais. Isso pode permitir que um órgão recrute e treine pessoas com conhecimentos essenciais em áreas como elaboração de provas, amostragem e análise estatística. Um baixo compromisso pode resultar num padrão de atribuir a avaliação nacional a diferentes organismos, uma estratégia que em nada contribui para criar a necessária competência técnica nas disciplinas relevantes. Em mais de um país, múltiplas agências têm realizado avaliações nacionais em separado, usando uma variedade de abordagens com valor limitado para a formulação de políticas de educação.

Em alguns casos, o compromisso do governo pode ser aumentado quando uma unidade dentro do ministério – com o apoio de uma dotação específica no orçamento para educação – realiza a avaliação. No Chile, por exemplo, o compromisso do governo e a resposta aos resultados do Sistema de Medição da Qualidade da Educação (SIMCE) aumentaram quando a avaliação nacional foi transferida de uma universidade para o ministério. A avaliação anual, a apresentação dos resultados num tempo hábil e o reconhecimento do valor dos resultados

para a formulação de políticas ajudaram a fortalecer a legitimidade do SIMCE, institucionalizar seu trabalho, garantir o compromisso adicional do governo e maior apoio. Em vários outros países latino-americanos, existem institutos de avaliação, independentes do Ministério da Educação, que conseguiram criar uma reputação de competência e autonomia, conduzindo avaliações com considerável flexibilidade e coerência (Ferrer, 2006).

A institucionalização em si não é suficiente, embora provavelmente contribua para evitar uma situação em que os resultados da avaliação nacional não cheguem às pessoas relevantes do governo. Também é necessário investir esforços na elaboração de procedimentos eficazes para comunicar os resultados aos interessados, dentro e fora do ministério. Além dos funcionários governamentais, as constatações da avaliação nacional são relevantes para o trabalho de elaboradores de currículos, comissões de exames, formadores de professores, e professores em sua prática diária nas escolas. Para atender às necessidades de informação de todos esses públicos, é necessário produzir vários relatórios e adotar diferentes estratégias de divulgação. As estratégias devem identificar os potenciais usuários (as principais instituições e os indivíduos-chaves) e seu nível de conhecimentos técnicos. É indispensável que exista um relatório técnico com informações suficientes para permitir a replicação do estudo. Mas os dados técnicos também precisam ser traduzidos em formas compreensíveis para usuários não técnicos, e podem ser apresentados num relatório sintético (por exemplo, para o público) ou num relatório mais detalhado para os formuladores de políticas. Esse relatório detalhado poderia indicar, por exemplo, se o sistema está deixando de atender plenamente a algum grupo em particular, se as lacunas encontradas requerem medidas corretivas e se os fatores associados ao desempenho superior podem ser identificados.

Em muitos países, a formulação de políticas tende a ser influenciada pelas prioridades políticas e pelas percepções dos ministros e de altos funcionários. Com frequência, o processo é desencadeado por experiências pessoais e informações esparsas, ou por pressões políticas; apenas muito raramente, baseia-se nos resultados de uma análise de dados válidos e confiáveis sobre o funcionamento do sistema educacional – como

seriam os resultados de uma avaliação nacional bem concebida e bem implementada.

Os formuladores de políticas devem assumir a liderança para garantir que as evidências objetivas e confiáveis sobre o funcionamento do sistema educacional, resultantes da avaliação nacional, sejam utilizadas para melhorar a qualidade global do processo de tomada de decisões. Podem fazer isso examinando resultados da avaliação nacional e refletindo sobre sua relevância para a elaboração de políticas em áreas como equidade de gênero e regional, fornecimento de materiais educativos nas escolas, qualificação dos professores e fornecimento de cursos durante o trabalho. Podem verificar se as mudanças introduzidas desde a avaliação nacional anterior parecem ter afetado o aproveitamento dos alunos. Podem incentivar e apoiar os cursos de formação de professores (cursos convencionais e cursos durante o trabalho), para que estudem os resultados e alterem as práticas atuais nos pontos em que as evidências revelam a necessidade de ajustes. Os formuladores de políticas também podem aconselhar a autoridade responsável a respeito de mudanças no conteúdo curricular quando a evidência claramente indicar que os alunos acharam o material muito fácil ou, mais provavelmente, muito difícil.

O envolvimento direto dos formuladores de políticas nos primeiros passos do projeto geral da avaliação e também na etapa posterior à conclusão, para discutir a relevância dos resultados, pode ajudar a garantir que venham a reconhecer o valor de uma avaliação nacional. Ao longo do tempo, pode-se esperar que os formuladores de políticas passem a considerar uma avaliação nacional instrumento-chave para o processo de formulação de políticas educacionais.

Breves descrições das práticas de avaliação nacionais em nove países são apresentadas no Apêndice A. As descrições não são exaustivas, e os casos não são apresentados como modelos perfeitos de boas práticas. Vários deles, na verdade, são falhos em muitos aspectos técnicos. No entanto, efetivamente revelam importantes semelhanças e diferenças na abordagem. As semelhanças têm a ver com o fato de que, em todos os países, as provas avaliaram o letramento em linguagem e em matemática em uma ou mais séries do ensino básico. Em todos os países, as ava-

liações foram baseadas em amostras. No Chile e no Uruguai, também foram feitas avaliações das quais participaram todas as escolas.

As diferenças entre os países referem-se à frequência com que são feitas as avaliações, que variou de um a quatro anos. O órgão responsável pela execução da avaliação também variou: o Ministério da Educação, um instituto de pesquisa apoiado pelo governo, ou uma comissão nacional de exames. Em diversos países, o órgão implementador contou com significativo apoio internacional. Em pelo menos dois (Chile e África do Sul), o órgão de implementação mudou entre uma avaliação e outra.

A forma de descrever o aproveitamento dos alunos também variou: a citação da média e da distribuição do número de itens aos quais os alunos responderam corretamente, a determinação do percentual de alunos cujo desempenho atingiu os padrões “esperados” ou o percentual da pontuação em diferentes níveis de “proficiência”. Os métodos de análise também variaram bastante, provavelmente como reflexo da capacidade técnica das equipes nacionais de avaliação. Sofisticadas abordagens analíticas foram utilizadas em alguns países (por exemplo, Estados Unidos e Vietnã).

A utilização dos resultados das avaliações pareceu variar bastante, embora não se possa ter certeza quanto a isso, porque é pouca a informação disponível em muitos países sobre em que medida os resultados foram divulgados ou efetivamente contribuíram para a formulação de políticas. Além de descrever as diferenças de gênero, alguns países utilizaram os resultados de uma avaliação nacional para apoiar as seguintes ações:

- Fornecer recomendações de políticas para o setor de educação (Sri Lanka, Vietnã).
- Documentar disparidades regionais no aproveitamento (Nepal, África do Sul, Sri Lanka).
- Estruturar um importante programa de treinamento de professores durante o trabalho (Uruguai).
- Proporcionar apoio financeiro e outros tipos de apoio às escolas com baixas pontuações (Chile).
- Levar ao conhecimento dos professores os pontos fracos e fortes no aproveitamento dos alunos (Uganda).

- Descrever, ao longo do tempo, as alterações no aproveitamento de alunos pertencentes a minorias (Estados Unidos).
- Sugerir redução da ênfase em álgebra e geometria no currículo (Butão).

As pessoas envolvidas no projeto de uma avaliação nacional talvez desejem conhecer uma série de práticas um tanto incomuns encontradas nas avaliações descritas no Apêndice A:

- Lançar uma campanha de sensibilização da opinião pública antes da avaliação (Chile).
- Coletar dados adicionais, simultaneamente à coleta dos dados sobre aproveitamento dos alunos, para monitorar em que medida as instalações da escola melhoraram ao longo do tempo (Vietnã).
- Aplicar o teste de desempenho aos professores, bem como aos alunos (Vietnã, Índia).
- Trabalhar em estreita colaboração com os sindicatos de professores para realizar a avaliação (Uruguai).

O Apêndice B fornece descrições das principais características dos três atuais estudos internacionais em larga escala que abrangem todo o mundo. Esses estudos concentram-se no letramento em três áreas: leitura, matemática e ciências. Essas três áreas de conhecimento e habilidades provavelmente seriam consideradas o “cerne” da formação dos alunos em todos os países. Os três estudos também se preocupam em acompanhar o aproveitamento dos alunos ao longo do tempo.

O nível de competência técnica em estudos internacionais é muito alto, e os países que participam deles podem, com isso, melhorar seus conhecimentos e habilidades. Muitos países, como vimos, também usam os dados coletados numa avaliação internacional para realizar análises em nível nacional, o que significa usar a avaliação internacional como uma avaliação nacional. Esse procedimento pode ser enriquecido se, além do requerido pelo estudo internacional, também forem colhidas informações contextuais específicas sobre o país em âmbito nacional.

O projeto dos estudos internacionais é muito semelhante ao de uma avaliação nacional, exceto quanto ao fato de que é preciso levar em con-

ta que a avaliação será realizada em vários países. Assim, os instrumentos de avaliação podem não ser igualmente adequados para todos os países, seja porque não representem bem os currículos das escolas, que podem variar de país para país, ou porque não reflitam toda a gama de aproveitamento dos alunos, que pode variar enormemente entre os países. Têm sido adotadas duas abordagens para tratar a variação nos currículos escolares. No TIMMS (B.1 no Apêndice B), bem como em estudos anteriores realizados sob os auspícios da Associação Internacional para Avaliação do Aproveitamento Escolar, as provas são desenvolvidas durante exercícios de construção de consenso entre os países participantes, e os elementos comuns de seus currículos são incluídos nas provas. A abordagem do Programa Internacional de Avaliação de Alunos (PISA) (B.3 no Apêndice B) tem sido não basear os instrumentos de avaliação numa análise de currículos, mas usar o parecer de “especialistas” para determinar os conhecimentos e habilidades que alunos de 15 anos deverão ter adquirido perto do final da escolaridade obrigatória, a fim de que possam participar plenamente da vida em sociedade.

O fato de que o aproveitamento dos alunos esteja relacionado com o desenvolvimento econômico dos países significa que avaliações concebidas para países industrializados (como o TIMSS e o PISA) provavelmente não fornecerão uma descrição satisfatória do aproveitamento num país em desenvolvimento. Foram criados estudos regionais para os países menos industrializados, a fim de resolver esse problema, e três desses estudos – dois na África e um na América Latina – estão descritos no Apêndice C. Esses estudos funcionaram como avaliações nacionais e também internacionais.



ESTUDOS DE CASO DE PAÍSES

A.1. ÍNDIA

Objetivo. Foi desenvolvida uma avaliação para ajudar o governo da Índia a fornecer dados de linha de base sobre a qualidade da educação em cada um dos estados. A avaliação é parte do programa governamental Sarva Shiksha Abhiyan (SSA), cujo objetivo era chegar a 2010 com a matrícula universal estendida a todo o ensino fundamental. Anteriormente, haviam sido realizadas avaliações de aproveitamento em larga escala em alguns distritos escolares como parte do programa governamental Educação Primária Distrital (Prakash, Gautam e Bansal, 2000). As pontuações médias em matemática e linguagem foram comparadas por distrito, área temática e nível de ensino. A avaliação concluiu que os alunos se saíram melhor em linguagem e que o aproveitamento médio na amostra de alunos mais velhos não foi tão impressionante quanto o dos alunos nas séries inferiores. A maior parte das diferenças entre os distritos, quando comparado o aproveitamento de meninos e meninas em matemática e linguagem, não foi estatisticamente significativa. Além dessa avaliação em nível distrital, realizou-se uma avaliação em grande escala em 22 estados no início de 1990 (Shukla *et al.*, 1994).

Frequência. A cada três anos.

Séries. A avaliação da 5ª série foi realizada em 2001-02. As 3ª e 4ª séries, bem como a última série do ensino fundamental (que varia de estado para estado), também foram avaliadas.

Aproveitamentos avaliados. Linguagem e matemática.

Quem fez? Conselho Nacional de Pesquisa e Treinamento (em Déli), com o apoio dos Institutos de Educação distritais, que supervisionaram a coleta de dados.

Amostra ou público-alvo. Amostra.

Análise. Foram apresentadas as pontuações da 5ª série em cada estado, em função do percentual de itens respondidos corretamente.

Utilização dos resultados. Os resultados da 5ª série mostraram pequenos hiatos nos níveis de aproveitamento nas comparações entre gêneros e entre escolas rurais e urbanas. Os dados serão utilizados para monitorar as mudanças nos níveis de aproveitamento escolar e para identificar fatores educacionais e não educacionais que podem ajudar a explicar as diferenças no aproveitamento dos alunos.

Pontos interessantes. Uma avaliação anterior em grande escala feita em 22 estados aplicou a mesma prova a professores e alunos. Em um estado em que os alunos tiveram pontuações médias muito baixas, apenas um dos 70 professores que realizaram a prova respondeu corretamente a todos os 40 itens de aritmética. Entre os professores, 10% responderam corretamente a menos da metade dos itens (Shukla *et al.*, 1994).

A avaliação nacional será usada para ajudar a monitorar o efeito da iniciativa SSA. Ao contrário do que ocorre na maioria das outras avaliações nacionais, as pontuações são divulgadas em função do percentual total de itens respondidos corretamente. Espera-se que os estados com níveis de aproveitamento particularmente insatisfatórios recebam aten-

ção especial. Alguns estados com forte tradição de ensino, em termos de elevadas taxas de escolarização (por exemplo, Kerala e Himachal Pradesh) registraram pontuações médias relativamente baixas na avaliação da 5ª série, enquanto alguns dos outros com taxas de escolarização relativamente baixas (por exemplo, Bihar, Orissa e Bengala Ocidental) tiveram notas mais altas. Esse resultado, que também foi encontrado numa avaliação anterior dos 22 estados, explica-se pelo fato de que, nos três últimos mencionados, as amostras de alunos que fizeram as provas tendiam a ser de “sobreviventes” no sistema educacional, pois muitos dos alunos menos favorecidos, em termos de contexto doméstico e níveis de habilidade, já haviam deixado a escola por volta da 5ª série.

Fonte: Índia, National Council of Educational Research and Training, Department of Educational Measurement and Evaluation (Conselho Nacional de Pesquisas Educacionais e Formação, Departamento de Mensuração e Avaliação Educacional), 2003.

A.2. VIETNÃ

Objetivo. Medir a qualidade da educação com enfoque particular no aproveitamento dos alunos no ensino básico.

Frequência. Avaliações anteriores em pequena escala foram realizadas entre 1998 e 2000 nas 3ª e 5ª séries, mas não forneciam informações adequadas que pudessem ser usadas como padrões de referência para monitorar as tendências ao longo do tempo.

Série. 5ª

Aproveitamentos avaliados. Leitura no idioma vietnamita e matemática em 2001.

Instrumentos. Testes de aproveitamento; questionários aplicados a alunos, professores e escolas.

Quem fez? Ministério da Educação e Formação, apoiado por outros órgãos nacionais, e uma equipe internacional apoiada pelo Banco Mundial e pelo Departamento para o Desenvolvimento Internacional do Reino Unido.

Amostra ou público-alvo. A amostra foi desenhada para ser representativa da população-alvo nacional e da população-alvo em cada uma das 61 províncias.

Análise. As análises incluíram tabulações cruzadas de dados de aproveitamento e dados das escolas por região, correlatos de aproveitamento, análise fatorial, modelagem de resposta ao item de dados sobre itens do teste, além de modelagem linear hierárquica para a identificação de fatores associados ao aproveitamento.

Utilização dos resultados. Funcionários governamentais fizeram 40 recomendações de políticas com base nos resultados gerais.

Pontos interessantes. As provas incluíram itens do Estudo sobre Letramento em Leitura realizado em 1991 pela Associação Internacional para Avaliação do Aproveitamento Escolar (Elley 1992, 1994) que foram usados para comparar os resultados com outros países. As mesmas provas foram aplicadas a professores e alunos; 12% dos alunos tiveram pontuações mais altas que as de 30% dos professores. Menos de 3% das escolas dispunham de recursos obrigatórios (por exemplo, biblioteca, água encanada). Mais de 80% dos alunos estavam em salas de aula que tinham recursos mínimos (quadro-negro, giz e assim por diante), enquanto 10% tinham professores que não haviam completado o ensino fundamental. Foram estabelecidos seis níveis de proficiência de acordo com o desempenho dos alunos na prova de leitura:

- *Nível 1.* Estabelece a correspondência entre palavras ou frases no texto com o auxílio de imagens. Restrito a um vocabulário limitado ligado a imagens.

- *Nível 2.* Localiza texto expressado em frases curtas e repetitivas e pode lidar com o texto sem a ajuda de imagens. O texto se restringe a sentenças curtas e frases com padrões repetitivos.
- *Nível 3.* Lê e compreende trechos mais longos. Pode procurar informação no texto, para trás ou para a frente. Compreende paráfrases. Um vocabulário mais amplo possibilita o entendimento de frases com estrutura relativamente complexa.
- *Nível 4.* Associa informações contidas em diferentes partes do texto. Seleciona e associa partes do texto para derivar e deduzir diferentes significados possíveis.
- *Nível 5.* Associa inferências e identifica a intenção de um autor a partir de informações expressas de diferentes modos, em diferentes tipos de texto e em documentos nos quais a informação não está explícita.
- *Nível 6.* Combina o texto e conhecimentos de fora para inferir vários significados, inclusive significados ocultos. Identifica os objetivos de um autor, atitudes, valores, crenças, motivos, pressupostos não declarados e argumentos.

Houve variação considerável no nível de desempenho dos alunos em ambas as provas de leitura e matemática. Por exemplo, o número de alunos que alcançaram os dois mais altos níveis de leitura foi muito menor em Ha Giang e Tien que em Da Nang (Tabela A.2.1). A relação entre as características do professor e as pontuações dos alunos foi estabelecida depois de levar em conta o contexto doméstico dos alunos (Tabela A.2.2).

Fonte: Banco Mundial 2004.

TABELA A.2.1**Porcentagens e Erros-Padrão de Alunos com Diferentes Níveis de Habilidade em Leitura**

| Província | Indicador de unidade | Nível 1 | Nível 2 | Nível 3 | Nível 4 | Nível 5 | Nível 6 |
|------------|----------------------|-------------|--------------|--------------|--------------|--------------|--------------|
| Ha Giang | Percentual EP | 7,5 1,66 | 22,1 3,23 | 27,4 3,06 | 18,7 2,97 | 18,5 3,07 | 5,7 2,09 |
| Tien Giang | Percentual EP | 2,8 0,7 | 13,4 2,0 | 28,8 2,49 | 20,2 1,8 | 22,4 2,46 | 12,5 2,78 |
| Da Nang | Percentual EP | 0,8 0,34 | 5,7 0,88 | 15,4 1,79 | 21,3 1,89 | 32,9 1,98 | 24,1 3,23 |
| Vietnã | Percentual EP | 4,6 0,17 | 14,4 0,28 | 23,1 0,34 | 20,2 0,27 | 24,5 0,39 | 13,1 0,41 |

Fonte: Banco Mundial, 2004, v. 2, Tabela 2.3.

Nota: EP = erro-padrão.

TABELA A.2.2**Relação entre Algumas Variáveis Relativas ao Professor e ao Aproveitamento em Matemática**

| Variável do professor | Correlação simples | Correlação parcial, depois de considerar o contexto doméstico do aluno |
|--|--------------------|--|
| Sexo do(da) professor(a) ^a | 0,17 | 0,14 |
| Formação acadêmica | 0,08 | 0,04 |
| Conhecimento de matemática | 0,29 | 0,25 |
| Classificado como "excelente professor(a)" | 0,18 | 0,13 |
| Recursos na sala de aula | 0,24 | 0,15 |
| Número de horas preparando e marcando | 0,00 | 0,01 |
| Frequência de reuniões com os pais | 0,05 | 0,04 |
| Número de visitas de inspeção | 0,13 | 0,11 |

Fonte: Banco Mundial, 2004, v. 2, Tabela 4.38.

Nota: As correlações acima de 0,02 são estatisticamente significativas.

(^a) Os alunos ensinados por professoras tiveram pontuações mais altas.

A.3. URUGUAI

Objetivo. A avaliação nacional destinava-se a identificar em que medida os alunos que concluíram o ensino básico haviam desenvolvido “entendimento fundamental” de linguagem e matemática, bem como os fatores socioculturais que podem ter influência sobre o aproveitamento dos alunos. A avaliação enfatizou o desenvolvimento profissional, que incluía o diagnóstico de problemas do aprendizado, dando aos professores informações sobre o desempenho dos alunos e ajudando-os a melhorar o ensino e a avaliação. A avaliação também pretendia utilizar os dados dos testes e questionários para melhorar as condições da escola.

Frequência e série. Sexta série (a cada três anos) em 1996, 1999, 2002 e 2005. Além disso, as 1ª, 2ª e 3ª séries foram avaliadas com propósitos de aperfeiçoamento dos professores em 2001. A 9ª série foi testada em 1999 e a 12ª, em 2003. Desde 2003, os alunos de 15 anos estão sendo avaliados como parte do Programa Internacional de Avaliação de Alunos (PISA).

Aproveitamentos avaliados. Matemática (resolução de problemas) e compreensão em leitura na 6ª série; matemática, linguagem, ciências naturais e ciências sociais nas 9ª e 12ª séries.

Instrumentos. Testes de aproveitamento; questionários para pais, professor e diretor da escola.

Quem fez? No início, a Unidade de Medição de Resultados Educacionais (UMRE), criada como parte de um projeto financiado pelo Banco Mundial, era a responsável pela avaliação nacional na 6ª série, enquanto o Programa de Modernização da Educação Secundária e Formação de Docentes (MESyFOD), um projeto financiado pelo Banco Interamericano, era responsável pela avaliação nacional no nível secundário. Desde 2001, as atividades de avaliação foram unificadas e institucionalizadas sob a Gerência de Pesquisa e Avaliação, parte da Administração Nacional da Educação Pública. O financiamento vem de agências internacionais.

Amostra ou público-alvo. Público-alvo e amostra de alunos da 6^a série, excluindo escolas rurais muito pequenas; todo o público-alvo da 9^a série; amostras das 1^a, 2^a, 3^a e 12^a séries; amostra para as avaliações do PISA.

Análise. A UMRE utilizou o corte de 60% de respostas corretas como um índice de adequação do desempenho dos alunos. As pontuações de cada escola foram comparadas com a média nacional, com a média distrital ou regional e com escolas que atendiam a alunos de origens socioeconômicas semelhantes. Os dados do teste de aproveitamento foram relacionados com fatores contextuais e com antecedentes.

Utilização dos resultados. Os resultados foram utilizados principalmente por professores, diretores e pela inspetoria escolar. O governo usou os resultados para identificar escolas que requeriam apoio especial e amplos programas de formação de professores durante o trabalho. Os resultados em nível nacional foram amplamente divulgados. Quarenta dias após a prova e antes do final do ano letivo, as escolas participantes receberam um relatório confidencial com os resultados agregados por escola, apresentando item por item. Os relatórios não incluíam os resultados individuais dos alunos nem os resultados desagregados por turma. A UMRE adotou várias providências: produziu manuais de ensino para ajudar a resolver as deficiências percebidas em linguagem e matemática e organizou programas de treinamento para professores durante o trabalho nas escolas situadas em áreas desfavorecidas; preparou relatórios para o pessoal de supervisão; organizou oficinas para os inspetores sobre questões relativas aos resultados da prova. As provas foram disponibilizadas para outras escolas que não integravam a amostra. Todas as escolas receberam um relatório das médias nacionais para cada competência testada. Escolas que não foram testadas receberam as normas nacionais para fins comparativos. Quase 80% dessas escolas aplicaram as provas e compararam seus resultados com as normas fornecidas. Os inspetores realizaram as próprias oficinas para desenvolver compreensão dos resultados, avaliar o efeito da privação social sobre os resultados do aprendizado do aluno e sugerir linhas de ação para melhorar a qualidade da educação.

Pontos interessantes. Inicialmente, o sindicato dos professores do primeiro grau fez forte oposição à avaliação nacional. Em particular, opunha-se à publicação dos resultados de cada escola. No final, o sindicato foi persuadido pelo compromisso do governo de não publicar os resultados de escolas ou professores, mas permitir que pudessem ser utilizados com fins de diagnóstico. Apenas dados agregados deveriam ser publicados. Além disso, o governo convidou os professores a participar dos grupos que planejarão a avaliação e de outros grupos consultivos. Os professores também tiveram grande envolvimento na elaboração das provas. Até o momento, não houve grande oposição a uma avaliação formal desse tipo no ensino básico. De modo geral, tem-se concordado que os professores ou escolas não podem ser penalizados por resultados ruins nos testes. O sindicato dos professores do ensino médio não tem sido muito favorável à avaliação e adotou a atitude de esperar para ver. A aceitação da iniciativa da UMRE e dos resultados pelos professores é atribuível à confidencialidade dos resultados das provas, à divulgação imediata dos resultados, à contextualização das pontuações de teste (levando em conta os dados do contexto sociocultural), ao reconhecimento de que os resultados dos alunos dependem de uma combinação de fatores (que incluem variáveis relativas a família, escola, comunidade e professores).

Embora os governos de alguns países estejam buscando modos de fazer com que as escolas e os professores também se responsabilizem pelos resultados escolares, o Uruguai adota uma abordagem diferente. O Estado assume a responsabilidade de promover um ambiente propício para ajudar a alcançar a igualdade no sistema educacional.

Fontes: Benveniste, 2000; Ravela, 2005.

A.4. ÁFRICA DO SUL

Objetivo. A África do Sul tem realizado uma série de avaliações nacionais nas 3ª, 6ª e 9ª séries. Também participou de três estudos internacionais que tinham como objetivo produzir dados de linha de base que possibilitassem acompanhar o progresso futuro e comparar os currículos

e o aproveitamento em matemática e ciências com os de países industrializados. Cada um desses estudos internacionais poderia ser considerado uma avaliação nacional do aproveitamento escolar. A participação numa avaliação internacional proporcionou oportunidade para o desenvolvimento de capacidades locais.

A África do Sul foi o único participante africano no TIMSS em 1995; no TIMSS de 1999, Marrocos e Tunísia também participaram, e, no TIMSS de 2003, agregaram-se a eles Botsuana, Gana e Egito. A África do Sul também participou da avaliação da 6ª série realizada em 2000 pelo Consórcio de Monitoramento da Qualidade da Educação da África Austral e Oriental (SACMEQ) e da avaliação da 4ª série (Monitoramento do Aproveitamento do Aprendizado) que começou em 1992.

Frequência. TIMSS, 1995, 1999 e 2003.

Série. 8ª

Instrumentos. Provas de aproveitamento; questionários aplicados a alunos, professores e diretores.

Aproveitamentos avaliados. Matemática e ciências.

Quem fez? Conselho de Pesquisas em Ciências Humanas em 1995 e 1999, e Universidade de Pretória em 2003.

Amostra ou público-alvo. Amostra. Uma turma completa da 8ª série foi examinada em cada escola selecionada.

Análise. O estudo comparou o desempenho dos alunos em matemática e ciências com o de outros países em função de desempenho médio e desempenho nos 5º, 25º, 50º, 75º e 95º percentis. Também comparou a África do Sul com outros países participantes em termos dos antecedentes e das atitudes dos alunos, currículos, características dos professores, características da sala de aula e os contextos de aprendizagem escolar e instrução; incluiu também uma comparação das pontuações médias de desempenho ao longo do tempo.

Utilização dos resultados. Os resultados do TIMSS têm servido de subsídio para debates parlamentares.

Pontos interessantes. A África do Sul tem 11 idiomas oficiais. Algumas palavras tiveram de ser traduzidas para o inglês sul-africano, e alguns contextos tiveram de ser modificados. Um tempo considerável foi dedicado a resolver problemas logísticos imputáveis à inadequação de alguns serviços, tais como e-mail e telefone, que em outros lugares estão amplamente disponíveis. A equipe de pesquisa nacional achou difícil cumprir os prazos impostos pelo TIMSS. O esforço inicial de amostragem revelou cerca de 4 mil escolas que não estavam no banco de dados nacional. A transferência de competências entre as equipes que realizaram as três avaliações TIMSS, no que se refere aos diversos aspectos envolvidos numa avaliação, foi limitada. Apenas um dos integrantes da equipe do primeiro TIMSS participou do TIMSS de 2003. A maioria dos alunos fez a prova escrita em uma língua diferente de sua língua de origem.

O segundo TIMSS foi utilizado para um estudo detalhado realizado no país (Howie, 2002). Os resultados incluíram o seguinte:

- As estatísticas oficiais indicavam turmas de tamanhos diferentes (muito maiores) do encontrado na amostra nacionalmente representativa de escolas participantes, o que sugere informações erradas sobre dados de matrículas escolares.
- Alguns alunos estavam com receio de que seu desempenho nas provas fosse contar em seu histórico escolar oficial. Alguns ficaram com medo de pedir ajuda. Muitos encontraram grande dificuldade com as questões abertas. Problemas adicionais foram causados por atrasos na chegada, absenteísmo e fraudes durante a aplicação da prova.
- Muitos alunos tiveram problemas para completar as provas e os questionários, por dificuldades de linguagem. Muitos professores não tinham a necessária fluência no idioma para se comunicar efetivamente com os alunos.
- Os professores gastaram muito tempo ensinando matérias que deveriam ter sido cobertas nas séries anteriores.

- Perto de 25% dos professores da 8ª série não tinham qualificação para ensinar matemática e não tinham formação de nível superior.
- Os alunos cujo idioma de origem era o inglês ou o africâner tiveram pontuações significativamente mais altas que as dos alunos que falavam outra língua africana em casa.
- Menos de 0,5% dos alunos atingiu o nível mais alto de desempenho em matemática, comparados a 10% da amostra internacional. A média (381) das pontuações mais altas alcançadas em nove províncias (Western Cape) foi significativamente inferior à pontuação média internacional do TIMSS (487).
- Nem a escola nem o tamanho da turma foram preditores significativos do aproveitamento em matemática.

Realizaram-se avaliações nacionais das 3ª, 6ª e 9ª séries solicitadas pelo Departamento de Educação, a fim de obter dados de linha de base para avaliações futuras e sugerir iniciativas de políticas. Cada uma dessas avaliações utilizou dados de questionários, bem como dados da prova de aproveitamento, para fornecer uma base a partir da qual seriam avaliados os esforços de longo prazo destinados a melhorar o acesso, a qualidade, a eficiência e a equidade. As comparações entre províncias produziram evidência de grandes diferenças regionais no aproveitamento. Os níveis gerais de desempenho foram considerados baixos. Por exemplo, a avaliação da 6ª série registrou percentual médio de pontuações corretas de apenas 38% em linguagem, 27% em matemática e 41% em ciências naturais. Foram elaborados relatórios específicos sobre a 6ª série para cada província, e o relatório nacional também teve uma versão específica sobre a 6ª série.

Fontes: Howie, 2000, 2002, Kanjee, 2006; Reddy, 2005, 2006.

A.5. SRI LANKA

Objetivo. Avaliar o aproveitamento dos alunos que haviam concluído a 4ª série em 2003.

Frequência. As avaliações anteriores haviam sido realizadas na 3ª série (1996) e na 5ª (1994, 1999). Além disso, outras avaliações foram realizadas na 4ª série (2007) e nas 8ª e 10ª séries (2005).

Série. 4ª

Aproveitamentos avaliados. Primeira língua (cingalês ou tâmil), matemática e inglês.

Instrumentos. Testes de aproveitamento; questionários aplicados a diretores, chefias seccionais, professores de classe e pais (veja Tabela A.5.1).

TABELA A.5.1

Dados sobre Antecedentes e Fonte na Avaliação Nacional do Sri Lanka

| Tipo de informação | Questionário | Seções | Número de questões |
|------------------------|---------------------|---|--------------------|
| Antecedentes da escola | Diretor | <ul style="list-style-type: none"> • Antecedentes gerais • Perfil do professor • Instalações da escola • Situação financeira • Opiniões | 37 |
| | Chefe de setor | <ul style="list-style-type: none"> • Antecedentes gerais • Instalações da escola • Procedimentos de avaliação do ensino-aprendizado • Opiniões | 13 |
| | Professor de classe | <ul style="list-style-type: none"> • Antecedentes gerais • Informações acadêmicas e profissionais • Detalhes da sala de aula • Opiniões | 41 |
| Família | Pais | <ul style="list-style-type: none"> • Antecedentes gerais • Instalações da casa • Situação socioeconômica • Apoio ao aprendizado • Opiniões | 51 |
| | Alunos | <ul style="list-style-type: none"> • Antecedentes gerais • Educação pré-escolar • Atividades pós-escola • Opiniões | 26 |

Fonte: Perera et al., 2004, Tabela 3.7.

Quem fez? Centro Nacional de Pesquisa e Avaliação em Educação, localizado na Faculdade de Educação da Universidade de Colombo.

Amostra ou público-alvo. Amostra desenhada para ser representativa da população nacional de alunos da 4ª série e da população dessa série em cada uma das nove províncias.

Análise. Foram feitas comparações das pontuações de aproveitamento por tipo de escola, localidade, gênero e nível de formação dos professores. As províncias e os distritos foram ordenados por sua classificação em cada área temática. Usou-se a detecção de relações causa-efeito (*path analysis*) para examinar as relações entre escola, contexto familiar e fatores relacionados com os alunos, por um lado, e aproveitamento dos alunos, por outro.

Utilização dos resultados. Os resultados foram usados para analisar o setor de educação, contribuir para o desenvolvimento de uma nova estratégia governamental e orientar o apoio dos doadores na área de educação. As constatações estão sendo utilizadas atualmente para estabelecer padrões de referência a partir dos quais se vem monitorando o nível de aproveitamento dos alunos em cada uma das províncias.

Pontos interessantes. A equipe da avaliação nacional do Sri Lanka selecionou a pontuação de 80% como ponto de corte para determinar o domínio em uma área.¹

Foram divulgadas as percentagens de alunos “com domínio” em cada uma das três áreas testadas. Os resultados sugerem que o padrão esperado foi fixado num nível exageradamente alto que não correspondia à realidade. Embora, com base nas pontuações médias, o relatório da avaliação tenha concluído que o desempenho global na primeira língua “parece ser de um padrão satisfatório” (Perera *et al.*, 2004, 47), quando o desempenho é avaliado com base no nível de domínio, surge um quadro

¹Aparentemente, essa decisão baseou-se num ponto de corte utilizado pela UNESCO em estudos anteriores de monitoramento do aproveitamento do aprendizado (UNESCO, 1990).

diferente. Menos de 40% dos alunos conseguiram o domínio na língua local e em matemática, e menos de 10% alcançaram esse nível em inglês. Os resultados mostraram grande disparidade de aproveitamento entre as províncias e os distritos (Tabela A.5.2). Foram identificados subgrupos com baixos níveis de aproveitamento. Relatórios específicos foram publicados para cada uma das nove províncias do país.

Fonte: Perera *et al.*, 2004.

A.6. NEPAL

Objetivo. A avaliação nacional de 2001 foi realizada para determinar em que medida o aproveitamento dos alunos havia mudado ao longo de um período de quatro anos, durante uma fase de grandes mudanças políticas.

Frequência. Dados de linha de base foram obtidos para alunos da 3ª série em 1997. (A 5ª série foi avaliada em 1999.)

Série. 3ª

Aproveitamentos avaliados. Matemática, nepalês e estudos sociais.

Instrumentos. Foram aplicados testes de aproveitamento em matemática, nepalês e estudos sociais a todos os alunos da amostra. Foram aplicados questionários a diretores e professores das três áreas temáticas em cada escola da amostra. Foram entrevistados 25% dos alunos e seus pais.

Quem fez: Centro de Serviços Educacionais e de Desenvolvimento.

Amostra ou público-alvo. Uma amostra de 171 escolas.

Análise. Pontuações de teste acima de 75% de respostas corretas foram classificadas como “desempenho satisfatório”. Outras análises incluíram estudos de confiabilidade de cada teste e comparações de pontuações

médias relativas a 1997 e 2001. Foi feita a análise de variância para comparar as pontuações médias de desempenho dos alunos entre regiões e adotou-se a análise de regressão múltipla para identificar fatores relacionados com o aproveitamento dos alunos.

Utilização dos resultados. Os resultados foram utilizados para monitorar mudanças no aproveitamento de 1997 a 2001 e, em particular, avaliar o efeito de mudanças em políticas, que incluíram aumento do orçamento, novos currículos, novos livros didáticos e materiais de ensino, novos centros de formação de professores. Foram identificadas as regiões com desempenhos mais altos. Em 2001, as diferenças entre as pontuações médias de meninos e meninas foram significativas apenas em matemática; os meninos tiveram pontuações médias mais elevadas. As pontuações médias gerais em estudos sociais foram significativamente mais altas em 2001 que em 1997.

Pontos interessantes. Os dados ajudaram a identificar áreas de currículo em que os alunos pareciam ter tido alguma dificuldade. Em matemática, os alunos foram capazes, de modo geral, de descrever palavras em números e números em palavras, medir o tempo e o peso, somar números em palavras até quatro dígitos e somar números decimais. Em geral, não conseguiam resolver problemas verbalmente usando qualquer das quatro operações básicas (adição, subtração, multiplicação, divisão). Em nepalês, o aluno médio tendia a ser capaz de ler uma história simples e usar algum vocabulário, mas não sabia ler corretamente nem responder a perguntas com base em passagens ou questões que descreviam uma história pictórica.

Os resultados da avaliação mostraram que muitas das reformas pareciam ter surtido pouco efeito. Mais de 60% dos professores disseram que suas aulas nunca haviam sido supervisionadas. Eles tendiam a receber relativamente pouco apoio durante o trabalho. Cerca de um terço não havia recebido treinamento. A instrução em sala de aula foi considerada ineficaz.

O relatório concluiu que, apesar das inúmeras reformas realizadas, provavelmente ainda era muito cedo para se observar melhoria no aproveitamento dos alunos. O relatório da avaliação nacional também des-

tacou a qualidade relativamente pobre do apoio à educação recebido em casa. Mais de 25% das mães eram analfabetas, e menos de 7% delas haviam concluído a 5ª série.

Fonte: Khaniya e Williams, 2004.

A.7. CHILE

Objetivo. O Sistema de Medição da Qualidade da Educação (SIMCE) do Chile foi originalmente concebido para ajudar a orientar os pais na escolha de uma escola. Ele agora busca três objetivos: fornecer informações ao Ministério da Educação sobre em que medida os alunos estão alcançando as metas de aprendizado consideradas mínimas pelo Ministério; fornecer informações aos pais, professores e às autoridades em nível municipal, regional e central; e fornecer aos formuladores de políticas dados que possam orientar a alocação de recursos para desenvolvimento de livros didáticos e currículos e para a formação de professores durante o trabalho, especialmente nas áreas mais carentes. O SIMCE busca melhorar o sistema educacional introduzindo procedimentos de avaliação que enfatizam avaliação, informação e incentivos. Também serve para sublinhar o compromisso do Ministério da Educação com a melhoria da qualidade e da equidade do sistema educacional.

O Chile também opera um sistema de avaliação à parte, mas relacionado ao SIMCE, para premiação da excelência no âmbito do SNED (Sistema Nacional de Avaliação do Desempenho Docente). O SNED abrange escolas públicas ou privadas que recebem financiamento estatal, fornecendo incentivos a professores e escolas para elevar os níveis de aproveitamento dos alunos.

Frequência. Anual.

Séries. 4ª e 8ª

Aproveitamentos avaliados. Espanhol (leitura e escrita), matemática, ciências naturais e ciências sociais.

Instrumentos. Os alunos fazem provas de aproveitamento, autoconceito e percepção. Questionários foram respondidos por diretores, professores e pais (apenas em um ano).

Quem fez? A primeira aplicação do SIMCE foi feita em 1978 por um órgão externo, a Pontifícia Universidade Católica do Chile. Atualmente, a avaliação é feita pelo Ministério da Educação.

Amostra ou público-alvo. Todos (praticamente todos) os alunos nas séries relevantes são avaliados em espanhol e matemática. Provas de ciências naturais, história e geografia são aplicadas a 10% dos alunos. Escolas muito pequenas em lugares inacessíveis são excluídas.

Análise. As escolas recebem uma classificação em comparação com outras escolas da mesma categoria socioeconômica, bem como uma classificação nacional. O SIMCE identifica as 900 escolas que compõem os 10% com as mais baixas pontuações nas provas de matemática e linguagem dentro de suas respectivas regiões/províncias e destina recursos especiais para elas (programa P-900).

Utilização dos resultados. Os resultados do SIMCE são amplamente utilizados em discussões sobre políticas. Seus relatórios informam, por sala de aula, o percentual médio de respostas corretas para cada objetivo avaliado, bem como o número médio de respostas corretas no total da prova. No início do ano escolar, o SIMCE divulga os resultados nacionalmente e também por escola, localização e região. Os manuais do SIMCE explicam os resultados e como os professores e as escolas podem usá-los para melhorar o aproveitamento dos alunos. As escolas beneficiadas pelo programa P-900 recebem apoio sob a forma de melhoria da infraestrutura, livros didáticos e bibliotecas para as salas de aula, material didático e oficinas de capacitação de professores durante o trabalho. As escolas são excluídas do programa P-900 quando suas pontuações no SIMCE excedem o ponto de corte de 10%.

O programa SNED usa as pontuações do SIMCE junto com quatro outras medidas de qualidade da escola. Os professores das escolas com

melhor desempenho dentro de uma região recebem uma premiação em dinheiro equivalente a um salário mensal, aproximadamente. Num esforço para assegurar a equidade, o ministério seleciona escolas que atendem a grupos socioeconômicos semelhantes, tanto urbanas quanto rurais e dos níveis básico ou fundamental. Embora diversos fatores sejam levados em conta no cálculo do índice, o aproveitamento escolar representa quase dois terços da pontuação do índice (Tabela A.7.1). O sistema de ponderação é periodicamente modificado para refletir as prioridades de políticas.

TABELA A.5.2
Porcentagem de Alunos que Alcançaram o Domínio da Primeira Língua, por Província

| Grupo | Classificação | Província | Percentual que alcançou o domínio | Percentual buscado |
|--------------|---------------|-----------------|-----------------------------------|--------------------|
| Acima de 50% | 1 | Ocidental | 53,5 | 80,0 |
| 26% a 50% | 2 | Meridional | 42,6 | 80,0 |
| | 3 | Ocidental Norte | 42,2 | 80,0 |
| | 4 | Sabaragamuwa | 40,2 | 80,0 |
| | 5 | Central Norte | 35,6 | 80,0 |
| | 6 | Uva | 33,9 | 80,0 |
| | 7 | Central | 33,8 | 80,0 |
| 1% a 25% | 8 | Oriental | 23,7 | 80,0 |
| | 9 | Norte-Oriental | 22,7 | 80,0 |

Fonte: Perera et al., 2004, Tabela 4.14.

Pontos interessantes. O SIMCE usa uma campanha maciça de relações públicas que inclui folhetos para pais e escolas, cartazes para as escolas, vídeos para oficinas, programas de televisão e comunicados de imprensa. Os relatórios são distribuídos para diretores, líderes municipais, supervisores escolares e funcionários de ministério. Os pais também recebem um relatório individualizado para sua escola. Os jornais publicam os resultados de cada escola. Como os municípios recebem financiamento do governo central em função do número de alunos, têm um interesse específico no resultado; bons resultados no SIMCE tendem a atrair mais alunos e, conseqüentemente, mais receita.

As escolas nas quais um grande número de alunos não comparece na data das provas não recebem os resultados. Algumas escolas superestimaram o nível de pobreza dos alunos para ajudar a aumentar suas chances de se qualificar para receber ajuda do programa P-900. A maior preocupação dos professores parece ser com a posição de sua escola em relação às escolas semelhantes, e não com a oportunidade de usar os resultados para promover o diálogo dentro da escola e ajudar a diagnosticar as áreas onde os alunos parecem ter maiores dificuldades. Alguns professores têm criticado a natureza excessivamente técnica dos relatórios sobre as escolas. O SIMCE dedica relativamente pouca atenção aos dados dos questionários aplicados a alunos, pais e professores. Constatou-se ser tecnicamente difícil medir as atitudes em face do aprendizado e dos valores dos alunos. O programa SNED presume que os incentivos financeiros inspirarão os professores a fazer maiores esforços para melhorar o aprendizado.

Fontes: Arregui e McLauchlan, 2005; Benveniste, 2000; Himmel, 1996, 1997; McMeekin, 2000; Olivares, 1996; Wolff, 1998.

TABELA A.7.1

Índice para Prêmios de Merecimento para Escolas no Chile, 1998-99

| Fator | Percentual |
|--|------------|
| Eficácia (pontuações SIMCE em matemática e ciências) | 37 |
| Valor agregado (ganho médio na pontuação SIMCE) | 28 |
| Iniciativa | 6 |
| Melhoria nas condições da escola | 2 |
| Igualdade de oportunidade | 22 |
| Cooperação pais-professor | 5 |

Fonte: Delannoy, 2000, Tabela 1.5.

A.8. ESTADOS UNIDOS

Objetivo. A Avaliação Nacional do Progresso Educacional (NAEP), que teve início em 1969, mede o aproveitamento escolar dos alunos e monitora mudanças no aproveitamento em idades e séries específicas. A NAEP, muitas vezes chamada de “The Nation’s Report Card”, também

examina aproveitamentos de subpopulações-alvo definidas por características demográficas e por antecedentes e contextos específicos. Na maioria dos estados, a amostra da NAEP é suficientemente grande para permitir inferências sobre o aproveitamento em cada estado.

Frequência. As avaliações de matemática e leitura são realizadas pelo menos uma vez a cada dois anos e com menor frequência em outras áreas de currículo.

Séries. 4^a, 8^a e 12^a. Avaliações estaduais específicas que utilizam as provas da NAEP limitam-se às 4^a e 8^a séries.

Aproveitamentos avaliados. Matemática, leitura, ciências, escrita, artes, cidadania, economia, geografia e história americana. Novas áreas temáticas que serão avaliadas: língua estrangeira e história geral.

Instrumentos. Testes de aproveitamento em leitura, matemática, ciências, escrita, história americana, cidadania, economia, geografia e artes. Um questionário (ao qual o aluno pode escolher responder ou não) no final do caderno de prova coleta informações sobre características demográficas dos alunos, experiências em sala de aula e apoio aos estudos em casa. Um questionário respondido pelo professor centra-se no histórico dos professores, sua formação e práticas de instrução. Um questionário relativo à escola busca informações sobre as políticas e características da escola. O professor fornece dados sobre alunos com deficiência ou que estão aprendendo o inglês.

Quem fez? Um Conselho Nacional de Avaliação, nomeado pelo secretário de Educação, tem a responsabilidade global pela NAEP. O conselho é composto por governadores, deputados estaduais, funcionários das escolas municipais e estaduais, educadores, representantes de empresas e membros do público em geral. Vários órgãos foram contratados para realizar segmentos específicos da NAEP. No período 2003-06, as atividades de desenvolvimento de itens, análise, amostragem e coleta de dados, distribuição e pontuação, e manutenção de páginas na internet estavam a cargo de diferentes órgãos.

Amostra ou público-alvo. Amostras das 4^a e 8^a séries em nível estadual (somente escolas públicas) e estudantes da 12^a série em nível nacional. O tamanho da amostra para cada teste da NAEP é de cerca de 2.500 alunos em cada estado. Um estudo em separado sobre as tendências de longo prazo divulga informações em nível nacional com os resultados em matemática e leitura para amostras de alunos com 9, 13 e 17 anos extraídas de escolas públicas e privadas.

Análise. Cada aluno responde apenas a uma parte do número total de itens do teste em determinada área de conteúdo. Os dados permitem comparações entre grupos (por exemplo, alunos do gênero masculino e do feminino em determinado estado). O modelo de resposta ao item é utilizado para estimar características da medição de cada questão da avaliação e criar uma escala única para representar o desempenho. São aplicados pesos às amostras para refletir as características do público-alvo. As escalas construídas permitem comparar avaliações realizadas em anos diferentes com públicos-alvos semelhantes. Medidas de controle de qualidade são aplicadas em cada fase da análise. São divulgados os percentuais de alunos em cada um dos três níveis de proficiência: “básico” (domínio parcial dos conhecimentos exigidos), “proficiente” (comando competente da matéria) e “avançado” (nível superior de desempenho).

Utilização dos resultados. Os resultados são amplamente divulgados. Porta-vozes políticos e outros têm utilizado os resultados da NAEP para expressar opiniões tanto positivas quanto negativas sobre a qualidade do sistema escolar americano.

Pontos interessantes. A NAEP monitora tendências no desempenho de subgrupos. Dá-se especial atenção às taxas de evolução dos subgrupos minoritários, principalmente ao aumento nas pontuações em leitura desde 1971. Em geral, as pontuações em leitura e matemática aumentaram entre os alunos da 4^a série, enquanto as diferenças de aproveitamento entre alunos de diferentes raças diminuíram. Em termos gerais, não foram registradas taxas de crescimento no aproveitamento em leitura durante um período em que duplicou o número de alunos latino-americanos (que, tradicional-

mente, têm tido dificuldade em dominar leitura em inglês). A mudança na natureza do público-alvo estudantil torna difícil determinar se os esforços para melhorar a pedagogia e o currículo estão tendo efeito.

Fonte: Johnson, 1992; Centro Nacional de Estatísticas de Educação (EUA), 2005, 2006.

A.9. UGANDA

Objetivo. A Avaliação Nacional do Progresso em Educação (NAPE) realizada em julho de 2005, no segundo semestre escolar, foi uma entre as diversas avaliações nacionais realizadas em Uganda. Os objetivos específicos da avaliação foram os seguintes:

- Determinar o nível de aproveitamento dos alunos em letramento em inglês e letramento em matemática.
- Analisar as relações entre aproveitamento, de um lado, e gênero e idade dos alunos, localização da escola (urbana, periférica, rural) e zonas do país, de outro.
- Examinar padrões de aproveitamento.
- Comparar aproveitamentos de alunos da 3ª e 6ª séries em 1999 e 2005.

Frequência. Uganda tem realizado avaliações nacionais de aproveitamento desde 1996. Inicialmente, pares de matérias (letramento em linguagem e em matemática; ciências e estudos sociais) eram avaliados a cada três anos. A partir de 2003, o foco tem sido nos dois tipos de letramento, que são avaliados anualmente.

Séries. 3ª e 6ª.

Aproveitamentos avaliados. Letramento em inglês e matemática. A fluência oral em inglês é avaliada a cada três anos.

Instrumentos. Testes de aproveitamento nos letramentos em linguagem e matemática. As avaliações nacionais anteriores usavam questionários

para alunos, professores e diretores. Avaliações que coletam dados com questionários são realizadas a cada três anos.

Quem fez? Comissão Nacional de Exames de Uganda (UNEB).

Amostra ou público-alvo. Inicialmente, foram amostrados distritos em cada uma das 14 zonas do país. O tamanho da amostra foi aumentado para garantir um mínimo de três escolas em cada distrito.

Análise. As pontuações dos alunos em cada teste foram classificadas em um dos quatro níveis: “avançado”, “adequado”, “básico” e “inadequado”. As pontuações correspondentes aos níveis foram determinadas e definidas quando os testes estavam sendo elaborados por painéis dos quais participavam funcionários do Centro Nacional de Desenvolvimento Curricular, faculdades de pedagogia que formavam professores do ensino básico, a Agência de Padrões Educacionais, a UNEB e profissões ligadas ao ensino. No teste de inglês de 50 itens aplicado na 3ª série, foram utilizados intervalos de pontuação para definir níveis de desempenho: 38-50, “avançado”; 20-37, “adequado”; 15-19, “básico”; e 0-14, “inadequado”. Os painéis decidiram que o nível adequado deveria ser considerado o nível mínimo “desejado” de proficiência. Menos de 40% dos alunos da 3ª série alcançaram o nível desejado de proficiência em inglês (Tabela A.9.1). Os resultados do teste de desempenho foram relatados (em termos percentuais) de acordo com idade dos alunos, localização da escola (urbana ou rural), região geográfica e zona.

TABELA A.9.1

Porcentagem de Alunos da 3ª Série em Uganda com Nível de Proficiência em Letramento em Inglês, 2005

| Classificação | Meninos (%) | Meninas (%) | Todos (%) |
|--|-------------|-------------|-----------|
| Proficiente (avançado + adequado) | 36,9 | 39,7 | 38,3 |
| Abaixo do nível desejado de proficiência (básico + inadequado) | 63,1 | 60,3 | 61,7 |

Fonte: UNEB, 2006, Tabela 3.02.

Utilização dos resultados. A UNEB imprimiu um cartaz para cada sala de aula das 3^a e 6^a séries em Uganda listando áreas de currículo em que o desempenho dos alunos em nível nacional foi considerado adequado (com dizeres como, por exemplo: “Podemos contar números”, ou “Podemos realizar adição e subtração de números escritos em números e símbolos) ou menos adequado (por exemplo, “Ajude-nos a desenvolver um vocabulário mais amplo” ou “Ajude-nos a fazer a divisão de números corretamente” ou ainda “Ajude-nos a resolver problemas verbais de matemática”). Cartazes semelhantes foram preparados para os professores.

A UNEB planeja disseminar as lições aprendidas a partir da avaliação nacional de 2005 sob a forma de relatórios separados, de uso simples, contendo as implicações da NAPE para professores, diretores, supervisores, inspetores, formadores de professores e formuladores de políticas. Também está projetando uma iniciativa-piloto para usar as abordagens da avaliação nacional a fim de melhorar a avaliação feita nas salas de aula.

Pontos interessantes. A maioria dos alunos teve de fazer os testes em sua segunda língua, pois seria muito difícil encontrar um idioma usado por todos. Mais de 25% das escolas do ensino básico não puderam ser incluídas na avaliação nacional devido, em parte, à agitação civil existente em determinadas regiões. A UNEB concluiu que algumas escolas haviam inflacionado o número de matrículas para obter mais recursos.

Muitos dos itens da prova de linguagem apareceram sob o título geral de “gramática” (50% para a 3^a série e 30% para a 6^a série). Em geral, os alunos acharam difíceis os itens do teste. Muitos estudantes tiveram pontuações relativamente baixas (veja Figura A.9.1). Embora o estudante típico da 3^a série devesse ter cerca de 8 a 9 anos, a idade média real dos alunos que fizeram o teste era 10,2 anos; alguns tinham 11 anos ou mais.

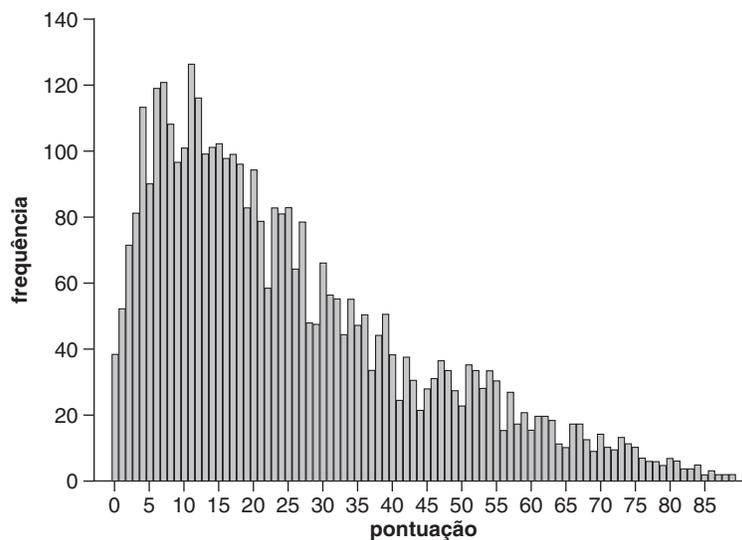
Foram encontradas diferenças significativas de aproveitamento entre as zonas. Na zona de Kampala, 87% dos estudantes da 6^a série alcançaram o nível desejado de proficiência em letramento em inglês; nas outras seis zonas, o percentual correspondente foi inferior a 30%. O desempe-

no no subteste de escrita da 6ª série revelou diferenças significativas entre os níveis de desempenho esperados e os reais. O nível desejado de proficiência foi alcançado por cerca da metade dos alunos que escreveram uma história sobre uma imagem, pela quarta parte dos que escreveram uma carta e por um décimo dos que compuseram e escreveram uma história. O relatório técnico incluiu uma amostra das cartas escritas pelos alunos e listas de erros comuns nos testes de matemática. Também incluiu inúmeras recomendações e uma lista de órgãos ou unidades que devem assumir a responsabilidade de dar seguimento a elas.

A UNEB recrutou os serviços de um consultor externo para fiscalizar a qualidade de seu trabalho, especificamente a qualidade das características estatísticas dos itens de prova e a correspondência entre os itens selecionados e os objetivos do currículo. O consultor observou uma estreita correspondência entre os itens e o currículo, mas recomendou que se desse maior atenção à resolução de problemas em matemática. O trabalho do

FIGURA A.9.1

Distribuição das Pontuações do Teste de Letramento na 6ª Série em Uganda



Fonte: Clarke, 2005.

consultor foi parcialmente limitado pela não disponibilidade de informações sobre as avaliações anteriores no que se referia à elaboração de provas, pesos da amostra, desenho do projeto e análise. Alguns dos problemas decorriam do fato de que uma parte dos trabalhos de análise da NAPE havia sido feita por um órgão não associado à UNEB. O consultor recomendou que cópias de todos os instrumentos, detalhes da amostragem, procedimentos analíticos e outras documentações relevantes sejam mantidas em arquivos pelo órgão nacional de avaliação (UNEB).

Fonte: UNEB, 2006.



ESTUDOS INTERNACIONAIS

B.1. TENDÊNCIAS INTERNACIONAIS NO ESTUDO DE MATEMÁTICA E CIÊNCIAS (TIMSS)

Marco de Referência

Os principais objetivos do TIMSS (Trends in International Mathematics and Science Study), organizado pela Associação Internacional para Avaliação do Aproveitamento Escolar (IEA), foram os seguintes:

- Avaliar o aproveitamento dos alunos em matemática e ciências, descrito em função de conceitos, processos, habilidades e atitudes.
- Descrever o contexto no qual se dá o aproveitamento dos alunos, a fim de identificar fatores associados ao aprendizado que pudessem ser melhorados por meio de mudanças em políticas (relativas, por exemplo, a ênfases no currículo, alocação de recursos ou práticas instrucionais).

Já foram realizados três estudos TIMSS. O primeiro cobriu 45 sistemas de educação em 1994-1995 e avaliou três públicos-alvo (3^a e 4^a séries; 7^a e 8^a séries; último ano do ensino fundamental). O segundo estudo cobriu 38 sistemas de educação em 1999 e avaliou alunos da 8^a

série. O terceiro foi realizado em 2003, cobrindo 50 sistemas e avaliando alunos das 4ª e 8ª séries. Estudos adicionais foram realizados em 2007 e 2008, e programados para 2011. O TIMSS distingue entre o currículo pretendido, o executado e o alcançado e, nas análises, explora a forma como estão inter-relacionados. O *currículo pretendido* representa uma declaração dos objetivos de ensino e aprendizado visados pela sociedade; esses objetivos são normalmente descritos em currículos, declarações de políticas e em regulamentos, e estão refletidos nos livros didáticos, recursos educacionais e exames. O *currículo implementado* é o modo como o currículo pretendido é interpretado pelos professores e oferecido aos alunos. Dados sobre a implementação (que fornecem um índice das oportunidades de aprender oferecidas aos alunos) são coletados principalmente por meio de questionários aplicados a professores e alunos. O *currículo alcançado* é o que os alunos aprenderam, conforme inferido de seu desempenho nos testes.

Instrumentos

Os seguintes componentes de matemática são avaliados nos testes do TIMSS:

- *Conteúdo*. Números; medição; geometria; proporcionalidade, funções, relações e equações; dados, probabilidade, estatísticas; análise elementar; validação e estrutura.
- *Expectativas de desempenho*. Conhecer, utilizar procedimentos de rotina, investigar e solucionar problemas, usar o raciocínio matemático, comunicar.
- *Perspectivas*. Atitudes, carreiras, participação, interesse crescente, hábitos da mente.

Os componentes de ciências do TIMSS compreendem o seguinte:

- *Conteúdo*. Ciências da terra, ciências biológicas, ciências físicas; ciência, tecnologia, matemática; história da ciência; questões ambientais; a natureza da ciência; a ciência e outras disciplinas.

- *Expectativas de desempenho.* Compreender; teorizar, analisar, resolver problemas; utilizar ferramentas, procedimentos de rotina e processos da ciência; investigar o mundo natural; comunicar.
- *Perspectivas.* Atitudes, carreiras, participação, interesse crescente, segurança, hábitos da mente.

Desde a criação do TIMSS, seus marcos de referência já foram alterados para refletir as mudanças curriculares e pedagógicas dos países participantes. Os elaboradores usaram uma grade curricular que se baseia em estudos anteriores (em particular, no caso da matemática, o Segundo Estudo Internacional de Matemática) para desenvolver testes por meio de um processo de construção de consenso entre os países participantes. Várias centenas de itens (de múltipla escolha e resposta fechada) passaram por um teste-piloto e foram avaliados em função de adequação e ajuste ao currículo. Para alcançar a cobertura máxima do currículo sem colocar um peso demasiadamente grande sobre os alunos que participaram do estudo, os itens do teste foram distribuídos por livretos. Cada aluno respondeu a apenas um livreto. A Tabela B.1.1 apresenta um exemplo retirado do documento-marco que orientou a avaliação feita pelo TIMSS em 2007.

Os questionários foram elaborados e aplicados para obter informações sobre o seguinte:

- Contextos sociais e educacionais gerais (*no nível do sistema*).
- Contextos local, da comunidade e das escolas (*no nível das escolas*).
- Aspectos dos antecedentes pessoais (*para cada aluno*).

Os instrumentos foram traduzidos em mais de 30 idiomas.

Participantes

Três populações-alvo participaram do TIMSS original em 1994-95:

- *Público-alvo 1.* Alunos do par de séries adjacentes que continham a maioria dos estudantes com 9 anos (geralmente as 3^a e 4^a séries).

TABELA B.1.1

Percentuais-alvos dos Testes de Matemática do TIMSS 2007 relativos aos Domínios de Conteúdo e Cognitivos, 4ª e 8ª Séries

| | | |
|--------------------------------------|--------------|----------|
| Domínios de Conteúdo na 4ª Série | Porcentagens | |
| Número | 50 | |
| Formas Geométricas e Medidas | 35 | |
| Apresentação de Dados | 15 | |
| Domínios de Conteúdo na Oitava Série | Porcentagens | |
| Número | 30 | |
| Álgebra | 30 | |
| Geometria | 20 | |
| Dados e Acaso | 20 | |
| Domínios Cognitivos | Porcentagens | |
| | 4ª Série | 8ª Série |
| Conhecimento | 40 | 35 |
| Aplicação | 40 | 40 |
| Raciocínio | 20 | 25 |

Fonte: Mullis et al., 2005, Mostra 2. Reproduzido com permissão.

- *Público-alvo 2.* Alunos do par de séries adjacentes que continham a maior parte dos estudantes com 13 anos (geralmente, as 7ª e 8ª séries).
- *Público-alvo 3.* Alunos do último ano do ensino fundamental. Foram identificadas duas subpopulações: (1) todos os estudantes que fizeram uma prova de letramento em matemática e leitura e (2) estudantes que estavam se especializando em matemática ou física ou que passaram por um teste especializado.

Em 1994-1995, 45 sistemas educacionais participaram do TIMSS (abrangendo os três tipos de público-alvo). Entre eles, um era do continente africano (África do Sul), seis estavam na Ásia/Oriente Médio (Hong Kong, China, Irã, Japão, Israel, Coreia do Sul; Kuwait, Cingapura e Tailândia) e um estava na região da América Latina e do Caribe (Colômbia). Os nomes dos sistemas educacionais que aparecem neste Apêndice são os listados nos relatórios dos estudos.

Em 1999, 38 sistemas educacionais participaram do TIMSS, cobrindo apenas o público-alvo 2. Entre eles, três estavam na África (Marrocos, África do Sul e Tunísia); 13 na Ásia/Oriente Médio (Taipei, Hong Kong, Indonésia, Irã, Israel, Japão, Jordânia, Coreia do Sul, Malásia, Filipinas, Cingapura, Tailândia e Turquia) e dois na América Latina e Caribe (Argentina e Chile).

Cinquenta países participaram do TIMSS 2003 (cobrindo os públicos-alvos 1 e 2). Seis estavam na África (Botsuana, Egito; Gana; Marrocos, África do Sul e Tunísia), 17 na Ásia/Oriente Médio (Bahrain; Taipei, Hong Kong, China; Indonésia, Irã, Japão, Israel, Jordânia, Coreia do Sul, Líbano, Malásia; Palestina, Filipinas, Arábia Saudita, Cingapura, Síria e Iêmen) e um na América Latina e no Caribe (Chile).

Alguns resultados

A Tabela B.1.2 apresenta os resultados do teste de matemática de 2003 para a 8ª série. Aproximadamente um terço dos alunos nos sistemas de mais alto desempenho alcançou o padrão de referência avançado. Em forte contraste, 19 dos sistemas com as pontuações mais baixas registraram no máximo 1% de alunos nesse nível de referência. Cingapura ficou em primeiro lugar tanto na prova da 4ª série quanto na da 8ª. Alguns sistemas demonstraram aproveitamento médio significativamente maior em relação aos seus desempenhos em 1995 e 1999, enquanto outros registraram quedas significativas na pontuação. Coreia do Sul, Hong Kong, China, Letônia, Lituânia e Estados Unidos estavam entre os que melhoraram as pontuações da 8ª série.

No conjunto, as diferenças entre gêneros no aproveitamento em matemática foram insignificantes. No entanto, as meninas superaram os meninos em alguns sistemas, enquanto os meninos se saíram melhor em outros. Um nível elevado de educação dos pais estava associado a pontuações de aproveitamento mais elevadas em praticamente todos os sistemas. No estudo de 2003, tanto na 4ª quanto na 8ª série o número de livros em casa teve correlação significativa com o aproveitamento dos alunos em matemática.

TABELA B.1.2

Distribuição do Aproveitamento em Matemática na 8ª Série no TIMSS

| Países | Anos de Escolaridade* | Média de Idade | Distribuição do Aproveitamento em Matemática | Escala de Pontuação Média | Índice de Desenvolvimento Humano **0,884 |
|---------------------------------------|-----------------------|----------------|--|---------------------------|--|
| Cingapura | 8 | 14,3 | | 605 (3,6) ● | — |
| ► Coreia do Sul | 8 | 14,6 | | 589 (2,2) ● | 0,879 |
| † Hong Kong | 8 | 14,4 | | 586 (3,3) ● | 0,889 |
| Taipei | 8 | 14,2 | | 585 (4,6) ● | — |
| Japão | 8 | 14,4 | | 570 (2,1) ● | 0,932 |
| Bélgica | 8 | 14,1 | | 537 (2,8) ● | 0,937 |
| † Países Baixos | 8 | 14,3 | | 536 (3,8) ● | 0,938 |
| Estônia | 8 | 15,2 | | 531 (3,0) ● | 0,833 |
| Hungria | 8 | 14,5 | | 529 (3,2) ● | 0,837 |
| Malásia | 8 | 14,3 | | 508 (4,1) ● | 0,790 |
| Letônia | 8 | 15,0 | | 508 (3,2) ● | 0,811 |
| Fed. Russa | 7 ou 8 | 14,2 | | 508 (3,7) ● | 0,779 |
| Rep. Eslovaca | 8 | 14,3 | | 508 (3,3) ● | 0,836 |
| Austrália | 8 ou 9 | 13,9 | | 505 (4,6) ● | 0,939 |
| † Estados Unidos | 8 | 14,2 | | 504 (3,3) ● | 0,937 |
| ‡ Lituânia | 8 | 14,9 | | 502 (2,5) ● | 0,824 |
| § Suécia | 8 | 14,9 | | 499 (2,6) ● | 0,941 |
| ¶ Escócia | 9 | 13,7 | | 498 (3,7) ● | 0,930 |
| ² Israel | 8 | 14,0 | | 496 (3,4) ● | 0,905 |
| Nova Zelândia | 8,5 - 9,5 | 14,1 | | 494 (5,3) ● | 0,917 |
| Eslovênia | 7 or 8 | 13,8 | | 493 (2,2) ● | 0,881 |
| Itália | 8 | 13,9 | | 484 (3,2) ● | 0,916 |
| Armênia | 8 | 14,9 | | 478 (3,0) ● | 0,729 |
| † Sérvia | 8 | 14,9 | | 477 (2,6) ● | — |
| Bulgária | 8 | 14,9 | | 476 (4,3) ● | 0,795 |
| România | 8 | 15,0 | | 475 (4,8) ● | 0,773 |
| Média Internacional | 8 | 14,5 | | 467 (0,5) | — |
| Noruega | 7 | 13,8 | | 461 (2,5) ● | 0,944 |
| Moldávia | 8 | 14,9 | | 460 (4,0) ● | 0,700 |
| Chipre | 8 | 13,8 | | 459 (1,7) ● | 0,891 |
| ² Macedônia | 8 | 14,6 | | 435 (3,5) ● | 0,784 |
| Libano | 8 | 14,6 | | 433 (3,1) ● | 0,752 |
| Jordânia | 8 | 13,9 | | 424 (4,1) ● | 0,743 |
| Irã | 8 | 14,4 | | 411 (2,4) ● | 0,719 |
| † Indonésia | 8 | 14,5 | | 411 (4,8) ● | 0,682 |
| Tunísia | 8 | 14,8 | | 410 (2,2) ● | 0,740 |
| Egito | 8 | 14,4 | | 406 (3,5) ● | 0,648 |
| Bahrain | 8 | 14,1 | | 401 (1,7) ● | 0,839 |
| Palestina | 8 | 14,1 | | 390 (3,1) ● | 0,731 |
| Chile | 8 | 14,2 | | 387 (3,3) ● | 0,831 |
| †‡ Marrocos | 8 | 15,2 | | 387 (2,5) ● | 0,606 |
| Filipinas | 8 | 14,8 | | 378 (5,2) ● | 0,751 |
| Botswana | 8 | 15,1 | | 366 (2,6) ● | 0,614 |
| Arábia Saudita | 8 | 14,1 | | 332 (4,6) ● | 0,769 |
| Gana | 8 | 15,5 | | 276 (4,7) ● | 0,567 |
| África do Sul | 8 | 15,1 | | 264 (5,5) ● | 0,684 |
| ¶ Inglaterra | 9 | 14,3 | | 498 (4,7) ● | 0,930 |
| Participantes do Padrão de Referência | | | | | |
| País Basco | 8 | 14,1 | | 487 (2,7) ● | — |
| Indiana (EUA) | 8 | 14,5 | | 508 (5,2) ● | — |
| Ontário (Prov. Canadá) | 8 | 13,8 | | 521 (3,1) ● | — |
| Quebec (Prov. Canadá) | 8 | 14,2 | | 543 (3,0) ● | — |

0 100 200 300 400 500 600 700 800

— Percentis de Desempenho —
5º 25º 35º 95º

Intervalo de Confiança de 95% para a Média (+25E)

● Média do país significativamente mais alta que a média internacional
⊙ Média do país significativamente mais baixa que a média internacional

- * Representa anos de escolaridade contados do primeiro ano do nível 1 da ISCED.
- ** Retirado do Relatório sobre Desenvolvimento Humano de 2003 do Programa das Nações Unidas para o Desenvolvimento, pp. 237-240.
- † Atendeu às diretrizes para taxas de participação na amostra somente depois que as escolas substitutas foram incluídas (Ver Anexo A.9).
- ‡ Quase atendeu às diretrizes para taxas de participação na amostra somente depois que as escolas substitutas foram incluídas (ver Anexo A.9).
- ¶ Não atendeu às diretrizes para taxas de participação na amostra (ver Anexo A.9).

- 1 População Nacional Desejada não cobre toda a População Internacional Desejada (ver Anexo A.6).
- 2 População Nacional Desejada cobre menos de 90% da População Internacional Desejada (ver Anexo A.6). A Coreia testou a mesma coorte de alunos que outros países, mas no final de 2003, quando começava o ano escolar seguinte.
- () Erros-padrão aparecem em parênteses. Como os resultados estão arredondados para o número inteiro mais próximo, alguns totais podem parecer inconsistentes.
- Um travessão (—) indica que dados comparáveis não estão disponíveis.

Fonte: Mullis et al., 2004, Anexo 1.1. Reproduzido com permissão.

A amplitude do currículo testado no TIMSS de 2003 variou entre os sistemas. Os relatórios dos professores sobre os alunos da 8ª série indicaram que, em média, 95% haviam recebido ensinamentos sobre números, 78% sobre medição, 69% sobre geometria, 66% sobre álgebra e 46% sobre dados. Mais de 80% dos alunos tiveram professores com pelo menos alguma formação profissional em matemática. Livros didáticos foram amplamente utilizados como base do ensino. O uso de calculadora, em contrapartida, variou muito entre os sistemas. O uso disseminado na 4ª série era permitido em apenas cinco sistemas. Escolas nas quais poucos alunos viviam em famílias economicamente desfavorecidas tiveram, em média, 57 pontos a mais na 8ª série e 47 pontos a mais na 4ª série que aquelas em que mais da metade dos alunos vinha de famílias desfavorecidas.

B.2. ESTUDO INTERNACIONAL SOBRE O PROGRESSO DO LETRAMENTO EM LEITURA (PIRLS)

Marco de Referência

O Estudo sobre Letramento em Leitura feito pela IEA em 1991 serviu de base para a definição de letramento em leitura do PIRLS (Progress in International Reading Literacy Study). Para o PIRLS (2001 e 2006), o letramento em leitura foi definido como

... a capacidade de compreender e utilizar as formas da linguagem escrita exigidas pela sociedade e/ou valorizadas pelo indivíduo. Leitores jovens podem construir significado a partir de uma variedade de textos. Eles leem para aprender, participar de comunidades de leitores e por prazer (IEA, 2000, 3).

O contexto de avaliação do PIRLS abrange dois grandes objetivos que foram combinados com quatro processos de compreensão. Os *objetivos* são os seguintes:

- *Literário*. Leitura para uma experiência literária na qual o leitor mergulha no texto para se envolver em eventos e personagens imaginados e para desfrutar o idioma.

- *Informativo*. Leitura para adquirir e utilizar informações, na qual o leitor se envolve com aspectos do mundo real representados em textos cronológicos (por exemplo, quando os eventos são descritos em biografias, receitas, instruções) ou em textos não cronológicos, em que as ideias são organizadas logicamente, e não cronologicamente (por exemplo, em textos de discussão ou persuasão).

Os *processos de compreensão* exigem dos alunos o seguinte:

- *Enfocar e recuperar informações explícitas*. Por exemplo, procurar ideias específicas; encontrar o tópico frasal ou ideia principal, quando explícita.
- *Fazer inferências simples*. Por exemplo, inferir que um evento causou outro; identificar generalizações no texto.
- *Interpretar e integrar ideias e informações*. Por exemplo, discernir a mensagem geral ou o tema de um texto; comparar e contrastar informações num texto.
- *Analisar e avaliar o conteúdo, a linguagem e elementos textuais*. Descrever como o autor concebeu um final surpreendente; avaliar a completude ou clareza da informação no texto.

O PIRLS foi realizado em 2001 e 2006.

Instrumentos

Estimou-se que o uso de textos “autênticos” (isto é, aqueles tipicamente lidos pelos alunos em suas experiências diárias) para cada um dos dois objetivos (leitura para experiência literária e leitura para adquirir e usar informação) exigiria quatro horas de teste. Dado que não parecia razoável esperar que qualquer aluno se sentasse por mais de uma hora numa situação de teste, o material da avaliação foi distribuído em 10 livretos, dos quais cada aluno respondeu a apenas um.

A habilidade dos alunos em cada um dos quatro processos de compreensão foi avaliada em questões que acompanhavam os textos. Foram utilizados dois formatos: múltipla escolha e resposta fechada.

Informações sobre as atitudes dos alunos com relação à leitura e sobre seus hábitos de leitura foram obtidas por meio de um questionário. Os

questionários também foram aplicados aos pais dos alunos, professores e diretores de escolas para coletar informações sobre o contexto familiar dos alunos e sobre experiências escolares consideradas relevantes para o desenvolvimento do letramento em leitura.

Participantes

O público-alvo do PIRLS foi definido como a série superior de duas séries adjacentes que tivessem o maior número de alunos com 9 anos. Na maioria dos sistemas, foi a 4ª série.

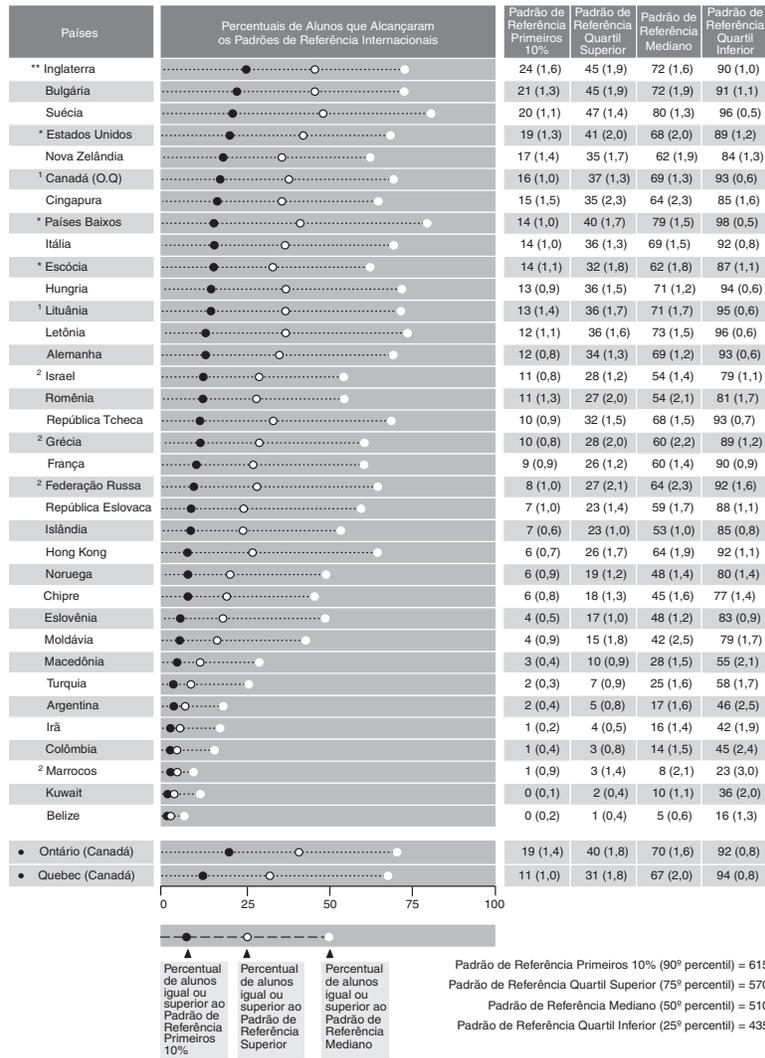
Trinta e cinco sistemas educacionais participaram do PIRLS em 2001. Incluíam um da África (Marrocos), seis da Ásia/Oriente Médio (Hong Kong, China, Irã, Israel, Kuwait, Cingapura e Turquia) e três da América Latina e do Caribe (Argentina, Belize e Colômbia) (Mullis *et al.*, 2003). Quarenta e um sistemas participaram do PIRLS em 2006. Entre os africanos, acrescentou-se a África do Sul. Na Ásia/Oriente Médio, entraram Taipei, Indonésia e Qatar, e saiu a Turquia. Na América Latina e no Caribe, saíram os três anteriores e entrou Trinidad e Tobago.

O próximo PIRLS está agendado para 2011.

Algumas Constatações

Foram criados quatro padrões de referências com base nas pontuações dos alunos no teste: o *quartil inferior de referência*, definido como o 25º percentil (o ponto acima do qual ficaram 75% dos alunos); a *mediana de referência*, definida como o 50º percentil; o *quartil superior de referência*, definido como o 75º percentil; e o *padrão superior de referência*, com os *10% de pontuações mais elevadas*, definido como o 90º percentil. Se as pontuações de aproveitamento em leitura tivessem a mesma distribuição em cada um dos países, cerca de 10% dos alunos em cada país seriam classificados no padrão de referência superior. A Tabela B.2.1 apresenta os resultados para os países participantes. Pode-se ver, por exemplo, que 24% dos alunos de inglês tiveram pontuações na categoria mais alta e que 10 sistemas tiveram menos de 5% dos alunos nessa categoria.

TABELA B.2.1
Percentuais de Alunos da 4ª Série que Alcançaram os Padrões de Referência
do PIRLS em Aproveitamento em Leitura



* O Canadá está representado apenas pelas províncias de Ontário e Quebec. A média internacional não inclui os resultados dessas províncias separadamente.
 † Atendeu às diretrizes para taxas de participação na amostra somente depois que as escolas substitutas foram incluídas (Ver Anexo A.7).
 ‡ Quase atendeu às diretrizes para taxas de participação na amostra depois que as escolas substitutas foram incluídas (ver Anexo A.7).
 ¶ A População Nacional Desejada não abrange toda a População Internacional Desejada. Como a cobertura

é inferior a 65%, o Canadá aparece como Canadá (O, Q), referindo-se apenas às províncias de Ontário e Quebec.
 2a A População Nacional Definida cobre menos de 95% da População Internacional Desejada (veja Anexo A.4)
 2b A População Nacional Definida cobre menos de 80% da População Internacional Desejada (veja Anexo A.4)
 () Erros-padrão aparecem em parênteses. Como os resultados estão arredondados para o número inteiro mais próximo, alguns totais podem parecer inconsistentes.

As meninas tiveram pontuações médias significativamente maiores que os meninos em todos os sistemas. Nos itens que mediam a leitura com fins informativos, os alunos na Suécia, nos Países Baixos e na Bulgária tiveram as maiores pontuações. Atividades precoces de alfabetização antes de iniciar a escola, como ler livros e contar histórias, estavam positivamente relacionadas ao desempenho posterior em leitura. Filhos de pais que tinham atitudes favoráveis à leitura tiveram pontuações mais elevadas no aproveitamento em leitura. Os alunos que falavam em casa a língua utilizada na avaliação tenderam a ter pontuações mais elevadas que os alunos que falavam outras línguas. As respostas dos diretores indicaram que, em todos os sistemas, e da 1ª à 5ª séries, a leitura era mais enfatizada que qualquer outra área dos currículos.

Na média, os professores disseram que a maioria dos alunos da 4ª série lia em voz alta para toda a turma diariamente. Usavam pouco as bibliotecas, embora essas tendessem a estar disponíveis. Em média, a maioria dos professores baseava-se nas próprias avaliações, e não em provas objetivas, para acompanhar o progresso do aluno. Quase dois em cada três alunos disseram que liam histórias ou romances pelo menos uma vez por semana. Em todos os sistemas, as atitudes dos alunos com relação à leitura estavam positivamente correlacionadas com o aproveitamento em leitura.

B.3. PROGRAMA INTERNACIONAL DE AVALIAÇÃO DE ALUNOS – PISA

Marco de Referência

O Programa Internacional de Avaliação de Alunos (PISA – Programme for International Student Assessment) avalia os conhecimentos e as habilidades dos alunos de 15 anos a cada três anos, sob os auspícios da Organização para a Cooperação e o Desenvolvimento Econômico (OCDE). O PISA foi desenvolvido para fornecer indicadores regulares de aproveitamento de alunos próximos do final do período de escolaridade obrigatória. Os dados destinam-se a alimentar os Indicadores Internacionais sobre Sistemas de Educação da OCDE.

Os alunos são avaliados em três domínios: leitura, matemática e ciências. Até o momento, foram realizadas três avaliações PISA. Em 2000, a leitura foi o domínio principal avaliado, ficando matemática e ciências como domínios secundários. Em 2003, a matemática foi o domínio principal e, em 2006, o foco ficou em ciências.

O PISA está projetado para ser usado por países individuais para (1) avaliar as habilidades de letramento dos alunos, em comparação com os alunos de outros países participantes, (2) estabelecer padrões de referência para a melhoria educacional, em comparação com o desempenho dos alunos de outros países, e (3) avaliar a capacidade nacional de fornecer elevados níveis de equidade nas oportunidades e nos resultados educacionais. O PISA tenta avaliar em que medida os alunos que se encontram perto do final do período de escolaridade obrigatória adquiriram alguns dos conhecimentos e habilidades essenciais para a plena participação na sociedade.

Participantes

Em 2000, 32 países participaram do PISA. Dois anos depois, outros 11 países se juntaram ao grupo. Nenhum país africano participou da avaliação de 2000. Entre os participantes da Ásia/Oriente Médio, estavam dois países da OCDE (Japão e Coreia do Sul) e cinco países “parceiros” não membros da OCDE (Hong Kong, Indonésia, Israel, Federação Russa e Tailândia). Na América Latina e no Caribe, participaram o México e também quatro países não membros da OCDE: Argentina, Brasil, Chile e Peru. Todos os 30 Estados-Membros da OCDE e mais 11 sistemas “parceiros” participaram em 2003. Entre os novos sistemas parceiros, havia um da África (Tunísia), um da Ásia (Macau) e um da América Latina e Caribe (Uruguai). Três sistemas parceiros originais (Argentina, Chile e Peru) não participaram da avaliação de 2003. E a Turquia, membro da OCDE, participou pela primeira vez em 2003. Em 2006, o número de sistemas participantes subiu para 57. A Tunísia continuou a ser o único participante africano. Novos sistemas parceiros na Ásia/Oriente Médio incluíam Azerbaijão, Taipei, Jordânia, Quirguistão e Qa-

tar. Todos os sistemas latino-americanos que participaram da avaliação em 2000 ou 2003 aplicaram as provas do PISA em 2006, que também foi aplicada pela Colômbia, um novo parceiro.

O público-alvo do PISA são alunos de 15 anos que integram uma amostra aleatória obtida em diferentes níveis de ensino nas escolas participantes.

Instrumentos

O teste de Letramento em Leitura pressupõe que os alunos sejam tecnicamente capazes de ler e busca avaliar sua capacidade de entender e refletir sobre uma vasta gama de materiais escritos em diferentes situações. São identificadas três dimensões: o *conteúdo* ou a *estrutura* dos textos (contínuos, como narrativa e descritiva, e não contínuos, como tabelas, gráficos e formulários); os *processos* que precisam ser executados (recuperação, interpretação, reflexão e avaliação); e a *situação* que gerou o conhecimento ou as habilidades, ou à qual são aplicados (pessoal, pública, ocupacional e educacional).

O teste de Letramento em Matemática diz respeito à capacidade dos alunos para analisar, raciocinar e comunicar ideias quando formulam, resolvem e interpretam problemas matemáticos em uma variedade de contextos. Foram destacadas três dimensões no contexto da matemática: *conteúdo* (espaço e forma, mudança e relações, quantidade e incerteza); *competências* (o grupamento de reprodução, o grupamento de conexões e o grupamento de reflexão) e *situações* (pessoais, educacionais ou profissionais, tanto públicas quanto científicas). Os itens do teste estão mais próximos das situações da “vida real” do que normalmente ocorre em provas convencionais de aproveitamento (veja Figura B.3.1)

O teste de Letramento Científico avalia a habilidade dos alunos de chegar a conclusões adequadas a partir das evidências e informações que recebem, criticar afirmações com base em evidências e distinguir entre opiniões e declarações baseadas em evidências.

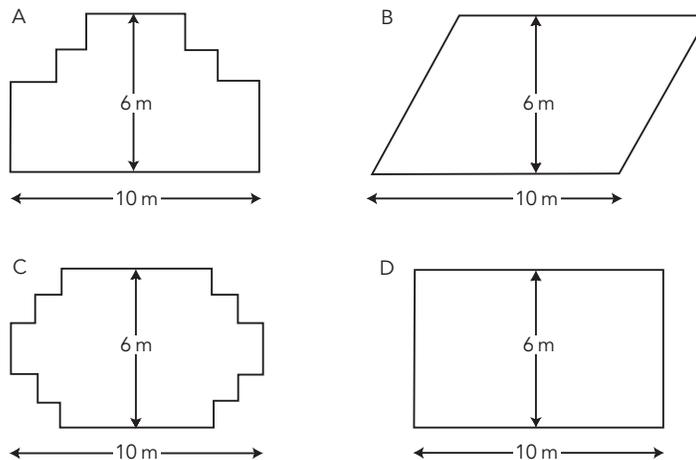
O marco de referência em ciências abrange três dimensões: *conceitos* científicos (selecionado da física, química, ciências biológicas e ciência

FIGURA B.3.1

Amostra de Itens do Teste de Matemática do PISA

CARPINTEIRO

Um carpinteiro tem 32m de madeira e quer fazer uma cerca em torno de um canteiro no jardim. Ele está pensando em usar um dos seguintes formatos:



| Pergunta 1 | |
|--|---|
| Faça um círculo em "Sim" ou "Não" em cada desenho para indicar se o canteiro pode ser feito com 32m de madeira | |
| Formato do canteiro | Usando este formato, pode-se fazer o canteiro com 32m de madeira? |
| Formato A | Sim / Não |
| Formato B | Sim / Não |
| Formato C | Sim / Não |
| Formato D | Sim / Não |

Fonte: OCDE, 2003. Reproduzido com permissão.

da Terra e do espaço); *processos* (descrever, explicar e prever fenômenos científicos; compreender a pesquisa científica); e *aplicação* (na vida e na saúde; na Terra e no meio ambiente; em tecnologia).

O fato de o número de itens de teste ser muito superior ao que poderia ser respondido por um aluno individual garante uma cobertura adequada dos domínios de interesse. Os itens de teste estão distribuídos em 13 livretos que contêm várias combinações de matemática, leitura, ciências e resolução de problemas.

Também foram aplicados questionários aos alunos (para obter informações sobre seu compromisso com o aprendizado, estratégias de aprendizado e crenças sobre si mesmos; sua percepção do ambiente de aprendizado e seus contextos familiares) e aos diretores das escolas (para obter informações sobre as políticas e práticas escolares e a qualidade dos recursos disponíveis) (OCDE, 2004b).

Algumas Constatações

O PISA divulga as pontuações médias dos países em uma tabela classificativa (Figura B.3.2). Também classifica o desempenho dos alunos por nível de proficiência. Para isso, toma como base aquilo que as pontuações de testes indicam que os alunos normalmente podem fazer. A Figura B.3.3 descreve as habilidades associadas a cada um dos seis níveis de proficiência do PISA de matemática. A Figura B.3.4 resume como os alunos de cada país se distribuíram entre os níveis de proficiência.

Os resultados indicam diferenças consideráveis entre países como Finlândia, Coreia do Sul e Canadá, onde a maioria dos alunos teve pontuação acima do Nível 2, e Brasil, Tunísia e Indonésia, onde apenas uma pequena minoria atingiu esse nível de proficiência. Outras constatações mostram que menos de 5% dos alunos nos países da OCDE alcançaram o Nível 6, enquanto cerca de um terço era capaz de desempenhar as funções associadas aos níveis 4, 5 e 6. Onze por cento dos alunos não foram capazes de executar as tarefas de matemática do Nível 1. Na maioria dos países, os meninos tenderam a obter pontuações mais elevadas que as meninas, especialmente em tarefas relacionadas a espaço e forma. Em alguns países (Austrália, Áustria, Japão, Países Baixos, Noruega e Polônia), as diferenças de aproveitamento entre gêneros não foram significativas. As meninas tenderam a ter menor interesse pela

FIGURA B.3.2

Pontuações Médias de Letramento em Leitura e Pontuações na Subescala de Leitura do PISA, 2000

| Pontuação combinada de letramento em leitura | | LEITURA NA SUBESCALA | | | | | |
|--|-------|----------------------------|-------|-------------------------|-------|-----------------------|-------|
| | Média | Recuperação de informações | | Interpretação de textos | | Reflexão sobre textos | |
| País | Média | País | Média | País | Média | País | Média |
| Finlândia | 546 | Finlândia | 556 | Finlândia | 555 | Canadá | 542 |
| Canadá | 534 | Austrália | 536 | Canadá | 532 | Reino Unido | 533 |
| Nova Zelândia | 529 | Nova Zelândia | 535 | Austrália | 527 | Irlanda | 533 |
| Austrália | 528 | Canadá | 530 | Irlanda | 526 | Finlândia | 533 |
| Irlanda | 527 | Coreia do Sul | 530 | Nova Zelândia | 526 | Japão | 530 |
| Coreia do Sul | 525 | Japão | 524 | Coreia do Sul | 525 | Nova Zelândia | 529 |
| Reino Unido | 523 | Irlanda | 523 | Suécia | 522 | Austrália | 526 |
| Japão | 522 | Reino Unido | 518 | Japão | 518 | Coreia do Sul | 526 |
| Suécia | 516 | Suécia | 516 | Islândia | 514 | Austria | 512 |
| Austria | 507 | Frância | 515 | Reino Unido | 514 | Suécia | 510 |
| Islândia | 507 | Bélgica | 515 | Bélgica | 508 | Est. Unidos | 507 |
| Reino Unido | 505 | Polónia | 505 | Austria | 508 | Polónia | 507 |
| Frância | 505 | Polónia | 502 | Eslovénia | 506 | Noruega | 506 |
| Noruega | 505 | Islândia | 500 | Eslovénia | 506 | Islândia | 501 |
| Est. Unidos | 504 | Est. Unidos | 499 | Est. Unidos | 505 | Dinamarca | 500 |
| Dinamarca | 497 | Suíça | 488 | Rep. Tcheca | 500 | Bélgica | 487 |
| Suíça | 484 | Dinamarca | 488 | Suíça | 496 | Frância | 486 |
| Espanha | 483 | Itália | 488 | Dinamarca | 494 | Grécia | 485 |
| Rep. Tcheca | 482 | Espanha | 483 | Espanha | 491 | Suíça | 488 |
| Itália | 487 | Alemanha | 483 | Itália | 489 | Rep. Tcheca | 485 |
| Alemanha | 484 | Rep. Tcheca | 481 | Alemanha | 488 | Itália | 483 |
| Hungria | 480 | Hungria | 478 | Polónia | 482 | Hungria | 481 |
| Polónia | 479 | Polónia | 475 | Hungria | 482 | Portugal | 480 |
| Grécia | 474 | Portugal | 485 | Polónia | 475 | México | 477 |
| Portugal | 470 | Grécia | 485 | Alemanha | 473 | Portugal | 477 |
| Luxemburgo | 441 | Luxemburgo | 450 | Polónia | 473 | Polónia | 477 |
| México | 422 | Luxemburgo | 433 | Luxemburgo | 446 | México | 446 |
| Média OCDE | 500 | México | 402 | Luxemburgo | 419 | México | 442 |
| | | Média OCDE | 488 | Média OCDE | 501 | Média OCDE | 502 |
| Países não OCDE | | | | | | | |
| Liechtenstein | 483 | Liechtenstein | 492 | Liechtenstein | 484 | Liechtenstein | 488 |
| Fed. Russa | 462 | Letónia | 451 | Fed. Russa | 468 | Letónia | 458 |
| Letónia | 468 | Fed. Russa | 451 | Letónia | 459 | Fed. Russa | 455 |
| Brasil | 386 | Brasil | 385 | Brasil | 400 | Brasil | 417 |

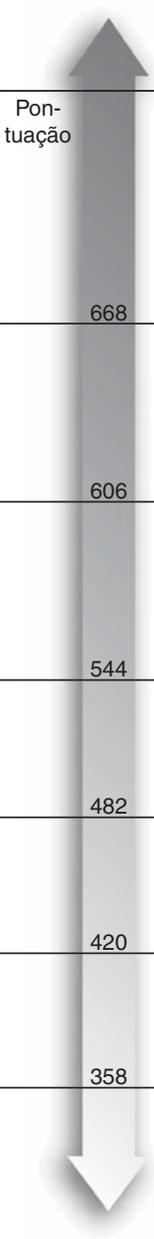
NOTA: Embora os Países Baixos tenham participado do Programa Internacional de Avaliação de Aunos (PIGA) em 2000, problemas técnicos com sua amostra impedem que os resultados sejam discutidos aqui. Para obter informações sobre os resultados dos Países Baixos, ver OCDE (2001). A média da OCDE é a média das médias nacionais de 27 países da OCDE. Como o PISA é principalmente um estudo da OCDE, os resultados para os países não membros da OCDE são apresentados separadamente dos países da OCDE e não estão incluídos na média da OCDE.

A média é significativamente superior à média dos EUA A média não é significativamente diferente da média dos EUA A média é significativamente inferior à média dos EUA

Fonte: OCDE, 2001, Figura 3. Reproduzido com permissão.

FIGURA B.3.3

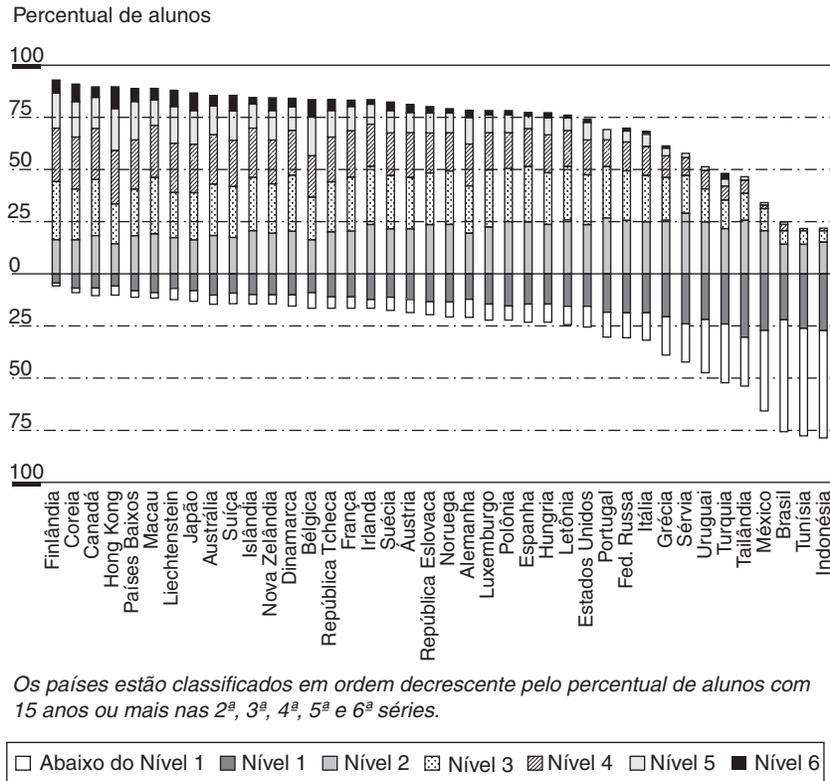
Níveis de Proficiência dos Alunos em Matemática no PISA

| | | |
|--|----------|--|
|  | | |
| Pon- tuação | Nível | O que os alunos podem fazer normalmente |
| 668 | 6 | No nível 6, os alunos podem conceituar, generalizar e utilizar informações com base em suas pesquisas e na modelagem de situações problemáticas complexas. Eles podem associar diferentes fontes de informação e representações, bem como transitar com flexibilidade entre elas. Os alunos neste nível são capazes de pensamentos e raciocínios matemáticos avançados. Esses alunos podem utilizar percepções e entendimentos, bem como domínio de operações e relações matemáticas simbólicas e formais, para desenvolver novas abordagens e estratégias em situações novas. Os alunos neste nível podem formular e comunicar com exatidão suas ações e reflexões a respeito das constatações, interpretações e argumentações, bem como a respeito da adequação dessas às situações originais. |
| 606 | 5 | No nível 5, os alunos podem desenvolver modelos para situações complexas e trabalhar com eles, identificando as limitações e especificando as hipóteses. Podem selecionar, comparar e avaliar as estratégias adequadas para lidar com problemas complexos relacionados a esses modelos. Os alunos neste nível podem trabalhar estrategicamente usando habilidades mentais e de raciocínio amplas e bem desenvolvidas, representações devidamente associadas ao caso, caracterizações simbólicas e formais e percepções relativas a essas situações. Podem refletir sobre suas ações e elaborar e comunicar interpretações e raciocínios. |
| 544 | 4 | No nível 4, os alunos podem trabalhar de forma eficaz com modelos explícitos para situações concretas complexas, e essas podem envolver restrições ou requerer que sejam feitas suposições. Podem selecionar e integrar diferentes representações, incluindo as simbólicas, associando-as diretamente a aspectos de situações reais. Os alunos neste nível podem utilizar habilidades bem desenvolvidas e raciocinar com flexibilidade, com algum discernimento, nesses contextos. Podem elaborar e comunicar explicações e argumentos com base em suas interpretações, ideias e ações. |
| 482 | 3 | No nível 3, os alunos podem executar procedimentos claramente descritos, inclusive aqueles que necessitam de decisões sequenciais. Podem selecionar e aplicar estratégias simples de resolução de problemas. Os alunos neste nível podem interpretar e utilizar representações baseadas em diferentes fontes de informação e raciocinar diretamente a partir delas. Podem elaborar comunicações breves relatando interpretações, resultados e seu raciocínio. |
| 420 | 2 | No nível 2, os alunos podem interpretar e reconhecer situações em contextos que requeiram não mais que uma inferência direta. Podem extrair informações relevantes a partir de uma única fonte e fazer uso de um único modo representacional. Os alunos neste nível podem empregar algoritmos básicos, fórmulas, procedimentos ou convenções. São capazes de raciocínio direto e de fazer interpretações literais dos resultados. |
| 358 | 1 | No nível 1, os alunos podem responder a questões que envolvem contextos conhecidos, em que todas as informações relevantes estão presentes e as perguntas estão claramente definidas. São capazes de identificar as informações e realizar procedimentos de rotina de acordo com instruções diretas em situações explícitas. Podem realizar ações óbvias e imediatamente derivadas do estímulo dado. |
| | | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 |

Fonte: OCDE, 2004a, Figura 1. Reproduzido com permissão.

FIGURA B.3.4

Percentual de Alunos em Cada Nível de Proficiência na Escala de Matemática do PISA



Fonte: OCDE, 2003b, Figura 2.16a. Reproduzido com permissão.

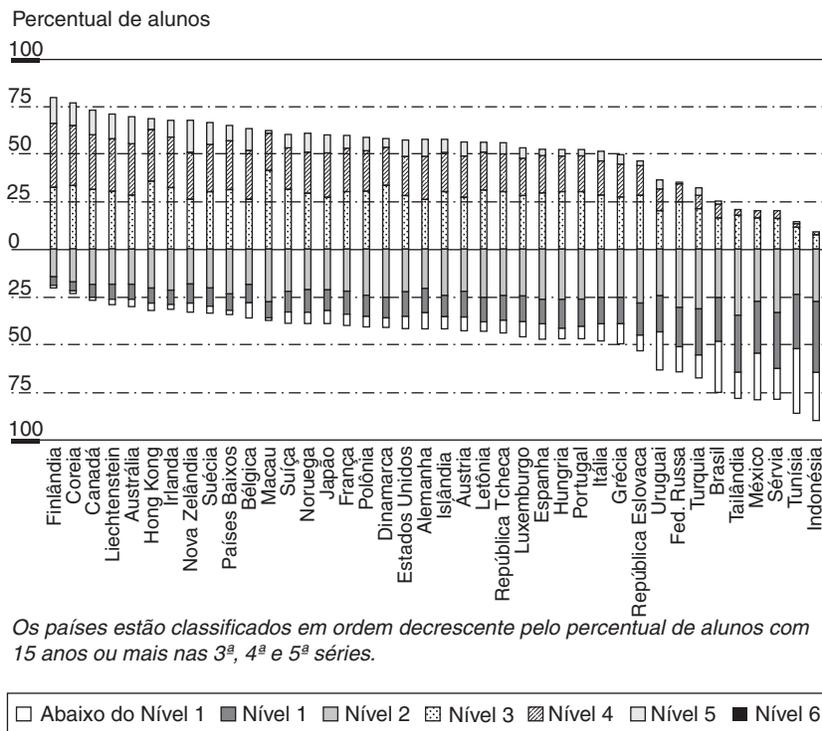
matemática, menor prazer com o tema, e alegaram ficar mais estressadas que os meninos nessa área de currículo. Os alunos americanos tenderam a ter “autoconceitos” mais fortes em matemática que alunos de outros países. Em contrapartida, os alunos do Japão e da Coreia do Sul, países que tiveram pontuações mais altas na prova de matemática, tenderam a ter autoconceitos relativamente fracos em matemática. A ocupação dos pais e o apoio à educação dos filhos mostraram forte correlação com o aproveitamento do aluno.

Raramente, observaram-se diferenças de gênero no aproveitamento em ciência: percentagens semelhantes de meninos e meninas tiveram

pontuações particularmente altas e baixas. Na leitura, a pontuação média da Finlândia ficou um nível e meio acima do nível de proficiência médio da OCDE. A Finlândia, assim como a Coreia do Sul e o Canadá, também registrou diferenças internas relativamente baixas, sugerindo níveis de equidade educacional mais elevados que os da maioria dos países participantes. Pouquíssimos alunos na Indonésia, Tunísia e Sérvia alcançaram o Nível 3 ou acima (veja Figura B.3.5).

FIGURA B.3.5

Percentual de Alunos em Cada Nível de Proficiência na Escala de Leitura do PISA



Fonte: OCDE, 2004b, Figura 6.2. Reproduzido com permissão.



ESTUDOS REGIONAIS

C.1. CONSÓRCIO DA ÁFRICA AUSTRAL E ORIENTAL PARA MONITORAMENTO DA QUALIDADE DA EDUCAÇÃO (SACMEQ)

Marco de Referência

O Consórcio da África Austral e Oriental para Monitoramento da Qualidade da Educação (SAMCEQ – Southern and Eastern Africa Consortium for Monitoring Educational Quality) é um agrupamento voluntário de ministérios da Educação da África Austral e Oriental que inclui Botsuana, Quênia, Lesoto, Malawi, Maurício, Moçambique, Namíbia, Seicheles, África do Sul, Suazilândia, Tanzânia, Zanzibar, Uganda, Zâmbia e Zimbábue. Lançado em 1995 com apoio do Instituto Internacional para o Planejamento Educacional (IIEP) da UNESCO, o SACMEQ foi concebido para (1) desenvolver a capacidade institucional por meio da formação conjunta (“aprender fazendo”, para planejadores educacionais) e da realização de pesquisas cooperativas sobre políticas educacionais que afetam a vida escolar e a qualidade da educação (por exemplo, identificando os pontos fracos dos sistemas educacionais em função de insumos e processos) e (2) monitorar mudanças no aproveitamento (IIEP, 2007). Uma característica notável do SACMEQ é sua estratégia sistemática de

consultar formuladores de políticas do alto escalão governamental para identificar questões de interesse que poderão ser abordadas em estudos empíricos. Também busca promover o envolvimento das partes interessadas e a maior transparência do processo de tomada de decisão. A primeira rodada de estudos SACMEQ foi realizada entre 1995 e 1999.

As preocupações de políticas que orientaram os estudos SACMEQ II realizados entre 2000 e 2003 foram agrupadas em cinco temas principais (Murimba, 2005b; Passos *et al.*, 2005):

- Características dos alunos e de seus ambientes de aprendizado.
- Características e percepções dos professores (por exemplo, sobre ensino e recursos).
- Características e percepções dos diretores de escolas (por exemplo, sobre o funcionamento das escolas e os problemas encontrados).
- Equidade na alocação de recursos humanos e materiais entre regiões e escolas.
- Aproveitamento em leitura e matemática dos alunos e de seus professores.

O SACMEQ baseou-se num estudo anterior (1991) realizado no Zimbábue (Ross e Postlethwaite, 1991) e começou como uma série de estudos nacionais. No entanto, tinha uma dimensão internacional, já que os estudos compartilhavam muitas características (questões de pesquisa, instrumentos, públicos-alvo, procedimentos de amostragem e análises). É feito um relatório separado para cada país. Foram feitas comparações transnacionais para o SACMEQ II, mas não para o SACMEQ I.

Instrumentos

Foi aplicado um teste de aproveitamento para coletar dados sobre os níveis de letramento dos alunos em leitura e matemática. Diversos itens do TIMSS foram incorporados às provas do SACMEQ II para se obterem dados comparativos. Utilizaram-se questionários para coletar dados de linha de base sobre insumos educacionais, condições gerais da vida

escolar e avaliações da equidade na distribuição de recursos humanos e materiais. Informações sobre o contexto familiar dos alunos foram obtidas por meio de questionários respondidos por eles; pediu-se aos alunos que marcassem numa lista os itens existentes em suas casas, como jornal diário, revista semanal ou mensal, rádio, televisão, telefone, motocicleta, bicicleta, água encanada e eletricidade.

Os testes do SACMEQ II incluíram itens selecionados de quatro estudos anteriores: Estudo dos Indicadores da Qualidade da Educação do Zimbábue, SACMEQ I, TIMSS e o Estudo sobre Letramento em Leitura realizado pela Associação Internacional para Avaliação do Aproveitamento Escolar (IEA). O uso desses itens tornou possível a comparação dos desempenhos registrados nesses estudos com o desempenho dos alunos testados no SACMEQ II.

Os relatórios dedicam espaço considerável à descrição das características dos professores (por exemplo, suas qualificações) e das condições nas escolas (por exemplo, mobiliário, material, tamanho e espaço da sala de aula), examinando como se comparam com os padrões de referência do ministério e como variam de acordo com a escola e a localização.

O SACMEQ II adotou a definição de letramento em leitura utilizada no Estudo sobre Letramento em Leitura da IEA (de 1990): “[A] capacidade de compreender e usar as formas da linguagem escrita exigidas pela sociedade e/ou valorizados pelo indivíduo” (Elley, 1992, 3). Também baseou o desenvolvimento do teste em três domínios identificados no estudo da IEA:

- *Prosa narrativa.* Texto contínuo em que o objetivo do autor é contar uma história, seja real ou ficcional.
- *Prosa expositiva.* Texto contínuo concebido para descrever, explicar ou transmitir informação factual ou opinião.
- *Documentos.* Informações estruturadas apresentadas sob a forma de gráficos, tabelas, mapas, listas ou conjuntos de instruções.

Construiu-se uma tabela de especificações em que os três domínios foram cruzados com sete níveis de habilidade em leitura:

- Lembrar literalmente
- Parafrasear conceitos
- Encontrar a ideia principal
- Fazer inferências a partir de textos
- Localizar informações
- Localizar e processar
- Aplicar regras.

O SACMEC II definiu letramento em matemática como “a capacidade de compreender e aplicar procedimentos matemáticos e tomar decisões a partir deles como indivíduo e membro da sociedade mais ampla” (Shabalala, 2005, 76). O teste avaliou a competência em três domínios:

- *Número*. Operações e linha de números, raízes quadradas, arredondamento e valor de lugar, números significativos, frações, porcentagens, razão.
- *Medição*. Relacionado com distância, comprimento, área, capacidade volumétrica, dinheiro, tempo.
- *Espaço-dados*. Formas geométricas, gráficos, tabelas de dados.

A tabela de especificações combinou esses três domínios com cinco níveis de habilidade “propostos” (ou esperados), que iam desde, por exemplo, realizar operações únicas simples utilizando números de até dois dígitos (Nível 1), até fazer cálculos envolvendo diversas etapas e uma mistura de operações com frações, decimais e números inteiros (Nível 5).

A maioria dos itens de teste estava no formato de múltipla escolha.

Os resultados foram apresentados em três formas: (1) pontuações médias, (2) porcentagens de alunos que atingiram os níveis mínimos e desejáveis de aproveitamento e (3) porcentagem de alunos que atingiram oito níveis de competência segundo o modelo que utiliza a técnica de escalonamento baseada na teoria da resposta ao item (Rasch).

As pontuações médias são medidas médias de desempenho e podem ser utilizadas para descrever o desempenho de diferentes categorias de

alunos (por exemplo, meninos e meninas, alunos que vivem em diferentes províncias e distritos).

Os níveis mínimos e desejáveis de aproveitamento foram definidos por comitês de especialistas (compostos por especialistas em currículo, pesquisadores e professores experientes) antes da coleta de dados. Foram identificados dois níveis:

- Um nível *mínimo*, indicativo de que um aluno mal conseguiria sobreviver durante o próximo ano de escolaridade.
- Um nível *desejável*, indicativo de que um aluno seria capaz de lidar com o próximo ano de escolaridade.

As análises foram realizadas para identificar os vários níveis de habilidades apresentados pelos alunos e proporcionar maior conhecimento sobre a natureza do aproveitamento. As habilidades de leitura associadas a oito níveis incluíram o seguinte:

- *Nível 1*. Pré-leitura: estabelece a correspondência entre palavras e imagens que envolvem conceitos concretos e objetos do cotidiano.
- *Nível 2*. Leitura emergente: estabelece a correspondência entre palavras e imagens que envolvem preposições e conceitos abstratos; usa sistemas de dicas para interpretar frases lendo o texto para diante.
- *Nível 3*. Leitura básica: interpreta o significado (fazendo a correspondência entre palavras e frases para completar uma frase) num texto curto e simples.
- *Nível 4*. Leitura de significado: lê para diante e para trás, a fim de associar e interpretar informações localizadas em várias partes de um texto.
- *Nível 5*. Leitura interpretativa: lê para diante e para trás, a fim de combinar e interpretar informações de várias partes de um texto em associação com informação externa (recordada) que completa e contextualiza o significado.
- *Nível 6*. Leitura inferencial: lê textos mais longos (narrativos, expositivos) para combinar informações de várias partes de um texto, a fim de inferir o propósito do autor.

- *Nível 7.* Leitura analítica: localiza informação em textos mais longos (narrativos, expositivos) para combinar informações e inferir as crenças pessoais do autor (sistema de valores, preconceitos, tendenciosidades).
- *Nível 8.* Leitura crítica: localiza informações em textos mais longos (narrativos, expositivos) para inferir e avaliar o que o escritor presumia sobre o tópico e sobre as características do leitor (por exemplo, idade, conhecimentos, crenças pessoais, valores).

As habilidades em matemática associadas a oito níveis incluíram o seguinte:

- *Nível 1.* Pré-letramento numérico: aplica operações de identificação ou subtração em uma única etapa; reconhece formas simples; encontra correspondência entre números e imagens; conta em números inteiros.
- *Nível 2.* Matemática emergente: aplica uma operação de soma ou subtração em duas etapas envolvendo execução e controle (por meio de estimativa básica); estima o comprimento de figuras familiares; reconhece formas bidimensionais comuns.
- *Nível 3.* Matemática básica: traduz a informação gráfica em frações; interpreta valor de lugar dos números inteiros até mil; interpreta unidades de medida simples do cotidiano.
- *Nível 4.* Início do letramento: usa operações matemáticas múltiplas em números inteiros, frações, decimais ou numa combinação de todos eles.
- *Nível 5.* Letramento competente: resolve problemas de operações múltiplas envolvendo unidades cotidianas de medição, números inteiros ou de outros tipos, ou todos eles.
- *Nível 6.* Matematicamente habilitado: resolve problemas de operações múltiplas envolvendo frações, razões e decimais; traduz informação representada verbal e graficamente em forma simbólica, algébrica e de equação.
- *Nível 7.* Resolução de problemas: extrai informações a partir de tabelas, gráficos e representações visuais e simbólicas para identificar e resolver problemas em várias etapas.

- *Nível 8.* Solução de problemas abstratos: identifica a natureza de um problema matemático não explícito, embutido em informações verbais ou gráficas, e traduz isso em forma algébrica ou numa equação, a fim de encontrar a solução.

Participantes

Entre 1995 e 1999, sete ministérios da Educação coletaram informações no SACMEQ I sobre letramento em leitura dos alunos da 6ª série. Entre 2000 e 2002, 14 ministérios completaram os estudos SACMEQ II como parte de um estudo sobre letramento em leitura e letramento em matemática dos alunos da 6ª série.

As condições variaram muito de país para país. Por exemplo, a renda nacional bruta das Seicheles (\$6.730) foi quase 40 vezes a do Malawi (\$170). Os gastos públicos em educação variaram entre 30% na Suazilândia e 7% na Tanzânia, enquanto o percentual de determinado grupo etário matriculado no ensino básico variou de cerca de 40% em Moçambique até pouco mais de 90% em Maurício, nas Seicheles e na África do Sul (Murimba, 2005b).

Tanto professores quanto alunos responderam aos testes de aproveitamento em vários países.

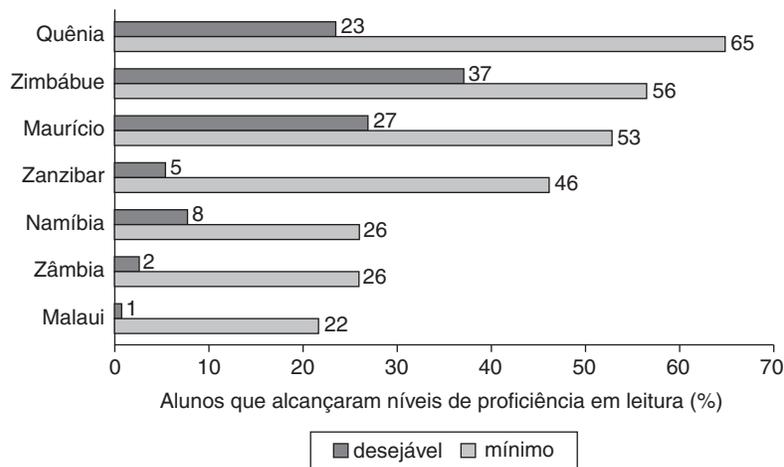
Alguns Resultados

Havia diferenças consideráveis de aproveitamento entre os países (Figura C.1.1)

Apenas 1% dos alunos da 6ª série no Malawi atingiu o nível “desejável” em leitura, enquanto no Zimbábue foram 37%. Quase 4 em cada 10 alunos nos países que participaram do SACMEQ II alcançaram o nível “mínimo” de domínio de leitura (definido por cada país antes da aplicação da prova), mas apenas 1 em 10 atingiu o nível “desejável”.

As comparações das pontuações de letramento em leitura de alunos em áreas urbanas e rurais revelaram grandes diferenças em favor dos

FIGURA C.1.1

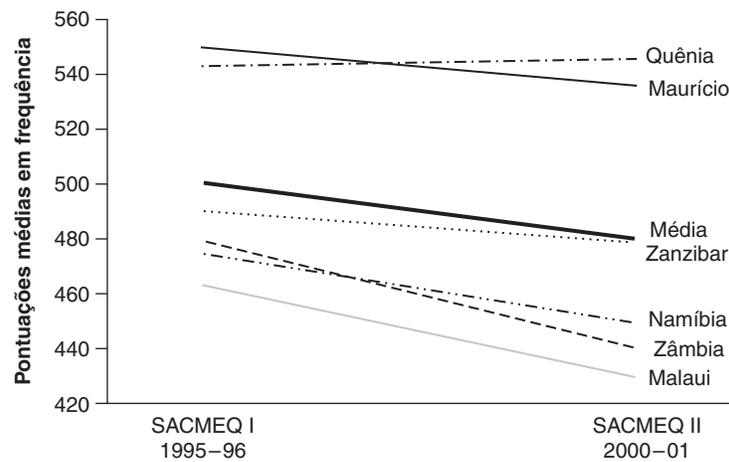
Percentual de Alunos da 6ª Série que Alcançaram Níveis de Proficiência em Leitura no SACMEQ, 1995-98

Fonte: UNESCO, 2004, Figura 3.1. Reproduzido com permissão.

alunos de áreas urbanas em quatro países (Quênia, Namíbia, Tanzânia e Zâmbia), enquanto em Maurício e nas Seicheles a diferença não foi estatisticamente significativa. As causas prováveis das diferenças urbano-rurais são complexas. Comparadas com alunos de áreas urbanas, os alunos de áreas rurais tinham famílias com níveis socioeconômicos mais baixos, eram mais velhos, tinham maior probabilidade de haver repetido um ano e recebiam menos apoio em casa para fazer o trabalho escolar. Além disso, as escolas rurais geralmente recebiam menos recursos (e de qualidade inferior) que as escolas urbanas, o que se refletia na forma como os professores atribuíam e corrigiam as tarefas de casa, na frequência com que se reuniam com os pais dos alunos e no nível de apoio dado pelos inspetores (Zhang, 2006).

Uma característica interessante do SACMEQ foi a utilização dos resultados para comparar a provisão de recursos e as tendências de aproveitamento em leitura num período marcado por rápido aumento nas matrículas escolares na região. Todos os seis sistemas educacionais que participaram do SACMEQ I (1995) e do SACMEQ II (2000) regis-

FIGURA C.1.2

Mudanças nas Pontuações de Letramento entre o SACMEQ I e o SACMEQ II


Fonte: UNESCO, 2004, Figura 2.4. Reproduzido com permissão.

traram aumento global no fornecimento de recursos às escolas entre as duas avaliações (Murimba, 2005a). Em cinco dos seis países, no entanto, as pontuações médias nacionais de letramento diminuíram (Figura C.1.2); essas diferenças foram estatisticamente significativas em apenas três países: Malauí, Namíbia e Zâmbia. Globalmente, as pontuações de aproveitamento caíram, em média, 4% nos seis países.

Cada relatório nacional produziu uma série de recomendações para os formuladores de políticas. Por exemplo, o relatório da Tanzânia recomenda que o governo investigue as disparidades de gênero na escolarização e identifique opções para ajudar a eliminá-las (Mrutu, Ponera e Nkumbi, 2005). Essa ação incluiria a prestação de cuidados a crianças órfãs para aliviar as meninas das pesadas responsabilidades domésticas que as impedem de frequentar a escola.

Alguns países também avaliaram os professores em função do domínio da disciplina usando a mesma prova aplicada aos alunos. Na Tanzânia, menos de metade dos professores atingiu o nível mais alto (Nível 8) em leitura (46,1%) ou em matemática (43,9%).

Os resultados do SACMEQ têm aparecido em comissões presidenciais e nacionais (no Zimbábue e na Namíbia), em revisões da política educacional por primeiros-ministros e membros do gabinete (em Zanzibar), em estudos setoriais sobre a educação nacional (na Zâmbia) e nas revisões de um plano diretor para a educação nacional (em Maurício).

Em vários países, os resultados foram interpretados como indicando a necessidade de se estabelecerem padrões para os recursos em educação. Por exemplo, o Quênia introduziu padrões de referência para o fornecimento de instalações de sala de aula (como mesas por aluno e livros por aluno). No Zimbábue, foram fornecidos fundos especiais para equipar as salas de aula.

Altas taxas de abandono e baixos índices de conclusão levaram o Ministério da Educação do Quênia a reforçar seu setor de educação não formal para atender aos que não se encaixam no sistema formal. Também no Quênia, as constatações do SACMEQ sobre gênero, disparidades regionais e ineficiências internas foram utilizadas para orientar o desenvolvimento de planos de ação, a fim de implementar o Educação Para Todos em nível nacional, provincial e distrital (Murimba, 2005a).

C.2. PROGRAMA DE ANÁLISE DOS SISTEMAS DE EDUCAÇÃO DA CONFEMEN – PASEC

Marco de Referência

O Programa de Análise dos Sistemas de Educação (PASEC – Programme d'analyse des Systemes Éducatifs de la Confemen) é realizado sob os auspícios da Conferência dos Ministros da Educação dos Países de Língua Francesa, ou CONFEMEN, na sigla em francês. Foi lançado em 1991, durante uma conferência de ministros da Educação de países francófonos em Djibuti, onde foi realizado em 1992 o primeiro estudo.

O PASEC tem como principal objetivo subsidiar a tomada de decisões em educação e, mais especificamente, tratar de questões importan-

tes da política educacional nacional. Para isso, avalia o aproveitamento dos alunos e busca identificar os fatores-chave associados a ele, bem como seus custos associados, a fim de estabelecer hierarquia do grau de eficiência de potenciais intervenções educativas.

Cinco características do PASEC são dignas de nota. Em primeiro lugar, tem uma dimensão internacional, na qual propostas de estudos nacionais são debatidas numa reunião dos países que integram a CONFEMEN. Caso a proposta seja aprovada, o representante nacional da CONFEMEN torna-se responsável pela criação de um grupo interdisciplinar de especialistas dentro do Ministério da Educação que, por sua vez, será responsável pela execução (projeto de questionário, aplicação, entrada e análise de dados, elaboração do relatório). O PASEC, no entanto, não foi concebido prioritariamente para comparar o aproveitamento dos alunos entre países.

Em segundo lugar, os alunos são testados no começo e no fim do ano letivo. Esse sistema significa que, nas análises, as características dos alunos no momento da entrada podem ser levadas em conta para se obter a medida de seu crescimento durante todo o ano.

Em terceiro lugar, estudos realizados em quatro países (Guiné, Mali, Níger e Togo) foram concebidos com um tema específico em mente. Por exemplo, Guiné e Togo adotaram como tema as políticas de emprego de professores (incluindo formação) introduzidas no Togo, em 1983, e na Guiné, em 1998, para reduzir o custo de contratação de mais professores – reconhecendo-se, ao mesmo tempo, que essas políticas poderiam afetar a qualidade da educação.

Em quarto lugar, a partir de 1995, os mesmos instrumentos foram utilizados em outros cinco países (Burkina Faso, Camarões, Costa do Marfim, Senegal em 1995-1996 e Madagascar, em 1997-1998), permitindo comparações internacionais.

Finalmente, em dois países (Costa do Marfim e Senegal), painéis representativos de alunos identificados na 2ª série em 1995 foram acompanhados por meio de estudos longitudinais até o final da 6ª série, em 2000.

Instrumentos

Foram construídos testes (de múltipla escolha e respostas fechadas) de francês e matemática com base em elementos comuns aos currículos dos países francófonos da África. Os testes foram projetados para aplicação no início e no final da 2ª e 5ª séries. As provas finais continham alguns itens das provas do início do ano, além de itens baseados no material abordado ao longo do ano. Na 2ª série, as provas de francês avaliaram o vocabulário de leitura dos alunos, a compreensão de sentenças e textos, e a escrita. Os testes da 5ª série, além de avaliar compreensão, analisaram também ortografia e aspectos de gramática.

As provas de matemática na 5ª série incluíram itens para avaliar o conhecimento dos alunos sobre as propriedades dos números e sua capacidade de realizar cálculos básicos (adição e subtração). Havia também itens nos quais os alunos deveriam usar adição, subtração, multiplicação e divisão para solucionar problemas, bem como itens para avaliar o conhecimento dos alunos sobre frações e decimais e conceitos geométricos básicos.

Também foram aplicados um teste de árabe em Maurício e um teste de malagasi em Madagascar. Em Camarões, uma tradução para o inglês do teste em francês foi aplicada a alunos de língua inglesa.

Os questionários aplicados aos alunos forneceram dados sobre suas características pessoais (gênero, idade, nutrição e língua falada) e sobre seus antecedentes e contextos (educação dos pais, disponibilidade de livros em casa e distância da escola); os questionários aplicados aos professores versaram sobre suas características pessoais (gênero, idade e educação ou formação) e o ambiente das salas de aula.

Nas análises, os dados sobre antecedentes e contextos foram cruzados com o aproveitamento dos alunos a fim de se identificarem possíveis relações entre os dois conjuntos de variáveis. Deu-se especial atenção ao “crescimento” ou ao “valor agregado” no decurso de um ano e à contribuição de fatores internos à escola (como nível de formação dos professores, tamanho da turma e disponibilidade de livros didáticos), bem como a fatores não escolares, como educação dos pais, distância da escola e língua falada em casa (Bernard, 1999; CONFEMEN 1999; Kulpo e Coustère, 1999).

Participantes

Até o momento, 18 países já participaram de atividades do PASEC: Benin, Burkina Faso, Camarões, República Centro-Africana, Chade, Costa do Marfim, República Democrática do Congo, Gabão, Djibuti, Guiné, Madagascar, Mali, Mauritânia, Maurício, Níger, República do Congo, Senegal e Togo.

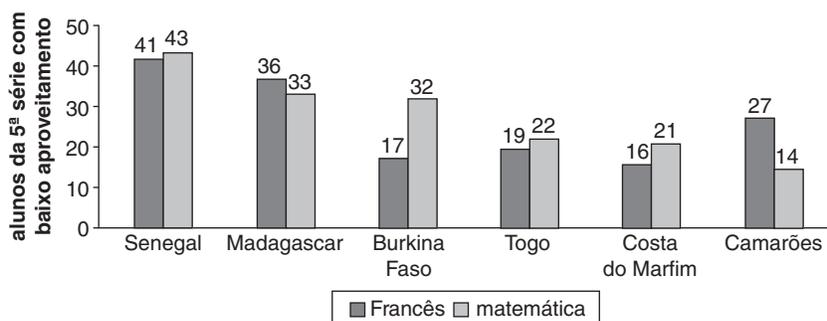
Algumas Constatações

Os resultados sugerem baixos níveis de aproveitamento, como os refletidos nas pontuações dos testes de leitura e matemática (Figura C.2.1). “Baixo aproveitamento” foi definido como uma pontuação abaixo do 25º percentil nas provas de leitura e matemática.

Já foram realizadas várias análises dos dados do PASEC. Numa delas, os dados de cinco países (Burkina Faso, Camarões, Costa do Marfim, Madagascar e Senegal) foram usados em um modelo linear hierárquico para avaliar as características do aluno, da escola e as de âmbito nacional

FIGURA C.2.1

Percentual de Alunos da 5ª Série com Baixo Aproveitamento, PASEC, 1996-2001



Fonte: UNESCO, 2004, Figura 3.32. Reproduzido com permissão.

Nota: A avaliação foi realizada em Burkina Faso, Camarões, Costa do Marfim e Senegal em 1995/96; em Madagascar, em 1997/98; e no Togo, em 2000/01. Os países estão classificados pela proporção de alunos com baixo aproveitamento em matemática. Baixo aproveitamento é definido como uma pontuação abaixo do 25º percentil em leitura e matemática.

que determinavam os aproveitamentos da 5ª série em francês e matemática (Michaelowa, 2001). Estas foram algumas das constatações que surgiram.

- Diversas características individuais dos alunos e das famílias (incluindo a alfabetização dos pais e o uso da língua francesa em casa) estavam relacionadas com o aproveitamento.
- Embora os alunos pudessem aparecer como beneficiários da repetência, os ganhos foram apenas temporários.
- Tanto a educação inicial dos professores quanto o treinamento regular durante o trabalho parecem fatores importantes na determinação do aproveitamento dos alunos.
- O número de dias em que os professores estiveram ausentes da escola afetou negativamente o aproveitamento dos alunos.
- Embora recebessem pagamento menor, os professores “voluntários” (empregados pelos pais dos alunos) eram mais eficazes do que os professores funcionários públicos.
- A filiação sindical dos professores tinha uma correlação estatisticamente significativa – e com sinal negativo – com o aproveitamento dos alunos.
- A disponibilidade de livros escolares teve um forte efeito positivo no aproveitamento do aprendizado.
- O tamanho da turma (até 62 alunos) estava positivamente relacionado com o aproveitamento.
- O aprendizado numa sala de aula multisseriada teve efeito positivo sobre o aproveitamento.
- Os alunos das escolas visitadas durante o ano por um inspetor apresentaram melhor desempenho que os das escolas não visitadas.
- O aproveitamento das meninas parecia se beneficiar do fato de terem uma professora, enquanto o dos meninos parecia melhorar quando tinham um professor.

C.3. LABORATÓRIO LATINO-AMERICANO DE AVALIAÇÃO DA QUALIDADE DA EDUCAÇÃO

Marco de Referência

O Primeiro Estudo Comparativo Internacional de Linguagem e Matemática na América Latina foi realizado pelo Laboratório Latino-Americano de Avaliação da Qualidade da Educação (LLECE – Latinoamericano de Evaluación de la Calidad de la Educación). Essa rede de sistemas nacionais de educação na América Latina e Caribe foi criada em 1994 e é coordenada pelo Escritório Regional da UNESCO para a América Latina e o Caribe.

O principal objetivo do estudo era fornecer informações sobre o aproveitamento dos alunos e os fatores associados que poderiam ser de utilidade para a formulação e a execução das políticas de educação nos países. Isso seria feito por meio da avaliação do aproveitamento das populações no ensino básico e abordaria as seguintes questões: O que os alunos aprendem? Em que série ocorre o aprendizado? Quais habilidades os alunos desenvolveram? Quando ocorre o aprendizado? Em que condições ocorre o aprendizado? (Casassus *et al.*, 1998).

A adoção de um marco comparativo foi vista como a melhor estratégia para ampliar a compreensão da situação educacional em cada país. Concluiu-se ser necessário realizar um estudo internacional na América Latina porque, além de poucos países da região terem participado de um estudo desse tipo, os instrumentos utilizados não levaram em conta as características específicas dos currículos da região.

Instrumentos

Foram elaborados dois testes de aproveitamento (em formulários separados), em linguagem e matemática, nos quais estava representado o conteúdo curricular de cada país participante. As provas foram de múltipla escolha, e a de linguagem também tinha questões abertas.

Os componentes de linguagem incluíam compreensão em leitura, práticas de metalinguística e produção de texto em espanhol – à exceção do Brasil, onde os alunos foram testados em português.

Os componentes de matemática incluíam números, operações com números naturais, frações comuns, geometria e medidas.

Recolheu-se ampla informação nos questionários (respondidos por alunos, professores, diretores e pais ou responsáveis) a respeito dos fatores que poderiam estar associados com o aproveitamento dos alunos (por exemplo, a localização e o tipo da escola, a escolaridade dos pais ou tutores e as percepções dos alunos e professores sobre a disponibilidade de recursos de aprendizagem na escola).

Participantes

Em 1997, 13 países participaram de uma pesquisa: Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, República Dominicana, Honduras, México, Paraguai, Peru e Venezuela. Os dados de 11 países estão incluídos no primeiro relatório da pesquisa.

Em cada país, foram avaliadas amostras de cerca de 4 mil alunos da 3ª série (com 8 e 9 anos) e da 4ª (com 9 e 10 anos). Os 20% “mais velhos da população total” foram excluídos (Casassus *et al.*, 1998, 18).

Alguns Resultados

Os resultados, classificados por tipo de escola (pública ou privada) e localização (cidades com mais de um milhão de habitantes, área urbana, área rural), indicam que os níveis de aproveitamentos dos alunos cubanos, qualquer que seja a localização da escola, estão muito à frente dos de outros países (Tabelas C.3.1 e C.3.2).

TABELA C.3.1
Percentual de Alunos em Cada Nível de Desempenho em Linguagem, por Tipo de Escola e Localização, LLECE 1997

| País | Pública | | | Privada | | | Megacidade | | | Urbana | | | Rural | | |
|-----------------|---------|----------|-----------|---------|----------|-----------|------------|----------|-----------|---------|----------|-----------|---------|----------|-----------|
| | Nível I | Nível II | Nível III | Nível I | Nível II | Nível III | Nível I | Nível II | Nível III | Nível I | Nível II | Nível III | Nível I | Nível II | Nível III |
| Argentina | 95 | 77 | 57 | 99 | 93 | 78 | 96 | 85 | 72 | 96 | 79 | 59 | 88 | 62 | 42 |
| Bolívia | 87 | 55 | 30 | 91 | 70 | 46 | 90 | 66 | 39 | 87 | 58 | 35 | 77 | 40 | 24 |
| Brasil | 95 | 80 | 54 | 98 | 93 | 72 | 96 | 88 | 62 | 95 | 82 | 58 | 84 | 62 | 38 |
| Chile | 93 | 71 | 49 | 97 | 86 | 67 | 94 | 76 | 53 | 95 | 79 | 60 | 89 | 63 | 41 |
| Colômbia | 89 | 59 | 35 | 97 | 81 | 56 | 96 | 79 | 53 | 89 | 60 | 36 | 89 | 57 | 33 |
| Cuba | 100 | 98 | 92 | n.a. | n.a. | n.a. | 100 | 99 | 93 | 100 | 98 | 92 | 100 | 98 | 92 |
| Rep. Dominicana | 77 | 52 | 30 | 83 | 64 | 42 | 84 | 65 | 42 | 73 | 44 | 25 | 73 | 39 | 20 |
| Honduras | 87 | 55 | 29 | 94 | 73 | 44 | 92 | 67 | 38 | 87 | 55 | 29 | 78 | 35 | 17 |
| México | 89 | 58 | 38 | 96 | 84 | 65 | 94 | 70 | 50 | 89 | 64 | 43 | 82 | 48 | 30 |
| Paraguai | 88 | 60 | 37 | 93 | 75 | 54 | n.a. | n.a. | n.a. | 90 | 67 | 44 | 81 | 51 | 32 |
| Peru | 86 | 55 | 29 | 94 | 78 | 54 | 92 | 70 | 43 | 85 | 57 | 34 | 71 | 30 | 13 |
| Venezuela | 88 | 59 | 38 | 91 | 70 | 49 | 91 | 68 | 48 | 88 | 60 | 38 | 84 | 58 | 39 |

Fonte: UNESCO, 2001, Tabela 8.

Nota: n.a. = não se aplica.

TABELA C.3.2
Percentual de Alunos em Cada Nível de Desempenho em Matemática, por Tipo de Escola e Localização, LLECE 1997

| País | Pública | | | Privada | | | Megacidade | | | Urbana | | | Rural | | |
|-----------------|---------|----------|-----------|---------|----------|-----------|------------|----------|-----------|---------|----------|-----------|---------|----------|-----------|
| | Nível I | Nível II | Nível III | Nível I | Nível II | Nível III | Nível I | Nível II | Nível III | Nível I | Nível II | Nível III | Nível I | Nível II | Nível III |
| Argentina | 96 | 54 | 12 | 98 | 71 | 23 | 98 | 70 | 26 | 96 | 54 | 11 | 94 | 43 | 6 |
| Bolívia | 93 | 43 | 9 | 96 | 59 | 18 | 95 | 49 | 12 | 94 | 51 | 14 | 89 | 36 | 8 |
| Brasil | 93 | 52 | 12 | 97 | 67 | 26 | 96 | 58 | 17 | 94 | 55 | 15 | 84 | 40 | 7 |
| Chile | 92 | 46 | 7 | 97 | 57 | 15 | 94 | 49 | 10 | 95 | 52 | 12 | 87 | 38 | 6 |
| Colômbia | 93 | 42 | 5 | 97 | 55 | 10 | 97 | 53 | 8 | 93 | 43 | 6 | 92 | 50 | 12 |
| Cuba | 100 | 92 | 79 | n.a. | n.a. | n.a. | 100 | 95 | 82 | 99 | 90 | 76 | 99 | 50 | 72 |
| Rep. Dominicana | 82 | 37 | 4 | 86 | 43 | 7 | 86 | 42 | 6 | 81 | 36 | 4 | 79 | 38 | 7 |
| Honduras | 84 | 36 | 7 | 93 | 39 | 5 | 87 | 35 | 3 | 86 | 39 | 8 | 78 | 23 | 13 |
| México | 94 | 55 | 10 | 98 | 69 | 20 | 97 | 62 | 13 | 94 | 58 | 13 | 90 | 46 | 10 |
| Paraguai | 87 | 29 | 2 | 90 | 49 | 12 | n.a. | n.a. | n.a. | 88 | 42 | 9 | 82 | 34 | 8 |
| Peru | 87 | 29 | 2 | 94 | 54 | 11 | 88 | 43 | 8 | 89 | 33 | 4 | 78 | 23 | 2 |
| Venezuela | 76 | 25 | 2 | 76 | 33 | 5 | 75 | 26 | 3 | 77 | 27 | 3 | 68 | 22 | 2 |

Fonte: UNESCO, 2001, Quadro 8.

Nota: n.a. = não se aplica.

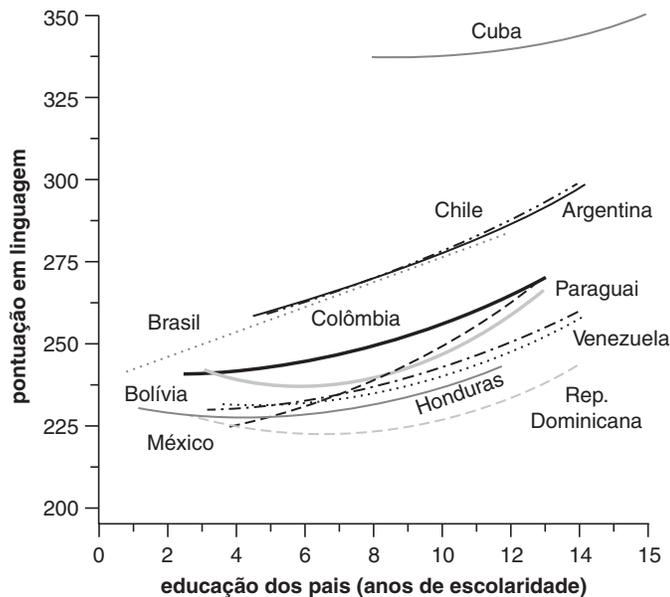
Mais de 90% dos alunos cubanos atingiram os níveis mais altos de proficiência (Nível III) em linguagem. Com uma exceção (escolas rurais), mais de 75% alcançaram esse nível em matemática. Enquanto 72% dos alunos das áreas rurais em Cuba alcançaram o Nível III em matemática, na maioria dos demais países esse percentual foi inferior a 10%.

Análises adicionais dos dados do LLECE centraram-se nas variações entre os países na relação entre situação socioeconômica (baseada no nível de escolaridade dos pais) e aproveitamento (veja Figura C.3.1).

Os dados indicam que os gradientes socioeconômicos variam consideravelmente entre os países; a relação é mais pronunciada na Argentina e no Brasil do que em Cuba, que teve relativamente pouca variação no nível de educação dos pais. Embora os alunos das escolas privadas tenham tido melhor desempenho que os das escolas públicas, as diferenças entre os grupos não foram significativas quando se levava em conta a situação socioeconômica do aluno (Cúpula das Américas, 2003).

FIGURA C.3.1

Percentual de Alunos em Cada Nível de Desempenho em Linguagem, por Tipo de Escola e Localização, LLECE 1997



Fonte: Willms e Somers, 2005.

Cuba teve a menor variação dos níveis de educação dos pais, bem como o nível mais alto de aproveitamento dos alunos. Outras análises revelaram que, em comparação com outros países, Cuba tendia a ter mais creches, mais atividades educacionais em casa, classes menores, mais professores altamente capacitados e menos salas de aula multisseriadas ou agrupadas por habilidades (Willms e Somers, 2001). Em um estudo de acompanhamento, os resultados do LLECE foram utilizados para identificar escolas com resultados excepcionais em sete países: Argentina, Bolívia, Chile, Colômbia, Costa Rica, Cuba e Venezuela (LLECE, 2002).

Apesar dessa variedade de análises, a Força-Tarefa para Reforma Educacional na América Central (2000, 19), em seu relatório intitulado “Tomorrow is too late” (Amanhã é tarde demais), observou que

em quase todos os casos, não existe uma política clara que estabeleça como os resultados da avaliação podem e devem ser usados. Testes de aproveitamento acadêmicos ainda não fazem parte das políticas de responsabilização sendo demandadas por vários grupos. Não tem havido discussões sobre o tipo de decisões que poderiam decorrer desses resultados, e há pouco consenso sobre o valor intrínseco de avaliar o desempenho dos alunos. Em consequência disso, esses programas são especialmente vulneráveis a mudanças no governo e até no pessoal que ocupa cargos elevados no ministério.



REFERÊNCIAS

- ARREGUI, P. e C. McLauchlan. 2005. "Utilization of Large-Scale Assessment Results in Latin America". Documento inédito preparado para o Partnership for Educational Revitalization in the Americas e o World Bank Institute.
- BANCO MUNDIAL. 2004. *Vietnam: Reading and Mathematics Assessment Study*. Vols. 1-3. Washington, DC: Banco Mundial.
- . 2007. Base de dados EdStats. <http://www1.worldbank.org/education/edstats/>.
- BEATON, A. E., T. N. Postlethwaite, K. N. Ross, D. Spearritt e R. M. Wolf. 1999. *The Benefits and Limitations of International Educational Achievement Studies*. Paris: UNESCO International Institute for Educational Planning.
- BENVENISTE, L. 2000. "Student Assessment as a Political Construction: The Case of Uruguay". *Education Policy Analysis Archives* 8 (32): 1-41.
- . 2002. "The Political Structuration of Assessment: Negotiating State Power and Legitimacy". *Comparative Education Review* 46: 89-118.
- BERNARD, J.-M. 1999. "Les Enseignants du Primaire dan Cinq Pays du Programme d'Analyse des Systèmes Educatifs de la CONFEMEN: Le Rôle du Maître dans le Processus d'Acquisition des Elèves". Relatório do Grupo de Trabalho sobre a Profissão de Docente, Seção Francófona, da Associação para o Desenvolvimento da Educação na África (ADEA). Paris: ADEA.
- BRAUN, H. e A. Kanjee. 2007. "Using Assessment to Improve Education in Developing Countries". In *Educating All Children: A Global Agenda* J. E. Cohen, D. E. Bloom e M. B. Malin (orgs.), 303-53. Cambridge, MA: MIT Press.

- BUTÃO, Câmara de Exames, Ministério da Educação. 2004. *National Educational Assessment in Bhutan: A Benchmark of Student Achievement in Literacy and Numeracy at Class 6, 2003*. Thimphu, Butão: Ministério da Educação.
- CAMPBELL, J. R., D. L. Kelly, I. V. S. Mullis, M. O. Martin e M. Sainsbury. 2001. *Framework and Specifications for PIRLS Assessment 2001*. 2ª ed. Chestnut Hill, MA: Boston College.
- CASASSUS, J., J. E. Froemel, J. C. Palafox e S. Cusato. 1998. *First International Comparative Study of Language, Mathematics, and Associated Factors in Third and Fourth Grades*. Santiago, Chile: Laboratório Latino-Americano de Avaliação da Qualidade da Educação.
- Centro Nacional de Estatísticas de Educação (EUA) 2005. *National Assessment of Educational Progress: The Nation's Report Card, Reading 2005*. U.S. National Center for Education Statistics, Washington, DC.
- . 2006. “NAEP Overview”. U.S. National Center for Education Statistics, Washington, DC. <http://nces.ed.gov/nationsreportcard/about/>.
- . s.d. “Comparing NAEP, TIMSS, and PISA in Mathematics and Science.” U.S. National Center for Education Statistics, Washington, DC. http://nces.ed.gov/timss/pdf/naep_timss_pisa_comp.pdf.
- CHINAPAH, V. 1997. *Handbook on Monitoring Learning Achievement: Towards Capacity Building*. Paris: UNESCO.
- CLARKE, M. 2005. *NAPE Technical Analysis and Recommendations*. Kampala: Uganda National Examinations Board.
- CONFEMEN 1999. *Les Facteurs de l'Efficacité dans l'Enseignement Primaire: Les Resultats du Programme PASEC sur Neuf Pays d'Afrique et de l'Océan Indien*. Dacar: CONFEMEN.
- Connecticut Department of Education. 2006. “State Releases Connecticut Mastery Test Results”. News, 9 de agosto. <http://www.sde.ct.gov/sde/lib/sde/PDF/PressRoom/2006cmtresults.pdf>.
- COULOMBE, S., J.-F. Tremblay e S. Marchand. 2004. *International Adult Literacy Survey: Literacy Scores, Human Capital, and Growth across Fourteen OECD Countries*. Ottawa: Estatísticas do Canadá.
- CRESPO, M., J. F. Soares, A. de Mello e Souza. 2000. “The Brazilian National Evaluation System of Basic Education: Context, Process, and Impact”. *Studies in Educational Evaluation* 26: 105–25.
- CÚPULA das Américas. 2003. *Regional Report: Achieving the Educational Goals*. Santiago: Ministério da Educação, Chile; Paris: UNESCO.
- Declaração Mundial sobre Educação para Todos*. 1990. Adotada pela Conferência Mundial de Educação para Todos: Atendendo Necessidades Básicas de Apre-

- dizado, Jomtien, Tailândia, 5–9 de março. New York: UNESCO. [Http://www.unesco.org/education/information/nfsunesco/pdf/JOMTIE_E.PDF](http://www.unesco.org/education/information/nfsunesco/pdf/JOMTIE_E.PDF).
- DELANNOY, F. 2000. *Education Reforms in Chile 1980–98: A Lesson in Pragmatism*. Washington, DC: Banco Mundial.
- EIVERS, E., G. Shiel, R. Perkins e J. Cosgrove. 2005. *The 2004 National Assessment of English Reading*. Dublin: Educational Research Centre.
- ELLEY, W. B. 1992. *How in the World Do Students Read? IEA Study of Reading Literacy*. Haia, Holanda: International Association for the Evaluation of Educational Achievement.
- (org.). 1994. *The IEA Study of Reading Literacy: Achievement and Instruction in Thirty-Two School Systems*. Oxford, Reino Unido: Pergamon.
- . 2005. “How TIMSS-R Contributed to Education in Eighteen Developing Countries”. *Prospects* 35 (2): 199–212.
- ETIÓPIA, National Organisation for Examinations. 2005. *Second National Learning Assessment of Ethiopia*. Addis Abeba: National Organisation for Examinations.
- FERRER, G. 2006. *Educational Assessment Systems in Latin America: Current Practice and Future Challenges*. Washington, DC: Partnership for Educational Revitalization in the Americas.
- FORÇA-TAREFA para a Reforma da Educação na América Central. 2000. *Tomorrow Is Too Late*. <http://thedialogue.org/publications/preal/tomorrow.pdf>.
- GANÁ, Ministério da Educação, Juventude e Esportes. 2004. *Results from Ghanaian Junior Secondary 2 Students' Participation in TIMSS 2003 in Mathematics and Science*. Acra: Ministério da Educação, Juventude e Esportes.
- GREANEY, V. e T. Kellaghan. 1996. *Monitoring the Learning Outcomes of Education Systems*. Washington, DC: Banco Mundial.
- HANUSHEK, E. A. e D. D. Kimko. 2000. “Schooling, Labor-Force Quality, and the Growth of Nations”. *American Economic Review* 90 (5): 1184–208.
- HANUSHEK, E. A., and L. Wössmann. 2007. *Education Quality and Economic Growth*. Washington, DC: Banco Mundial.
- HIMMEL, E. 1996. “National Assessment in Chile”. In *National Assessments: Testing the System*, org. P. Murphy, V. Greaney, M. E. Lockheed e C. Rojas, 111–28. Washington, DC: Banco Mundial.
- . 1997. “Impacto Social de los Sistemas de Evaluación del Rendimiento Escolar: El Caso de Chile”. In *Evaluación y reforma educativa: Opciones de política* (org.). B. Álvarez H. e M. Ruiz-Casares, 125–57. Washington, DC: ABEL/PREAL/AID.
- HORN, R., L. Wolff e E. Velez. 1992. “Educational Assessment Systems in Latin America: A Review of Issues and Recent Experience”. *Major Project of Education in Latin America and the Caribbean Bulletin* 27: 7–27.

- HOWIE, S. 2000. "TIMSS-R in South Africa: A Developing Country Perspective". Texto apresentado durante o encontro anual da American Educational Research Association, Nova Orleans, 24–28 de abril.
- . 2002. "English Proficiency and Contextual Factors Influencing Mathematics Achievement of Secondary School Pupils in South Africa". Tese de Doutorado, Universidade de Twente, Holanda.
- HOWIE, S. e C. Hughes. 2000. "South Africa". In *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science*, org. D. Robitaille, A. Beaton e T. Plomp, 139–45. Vancouver, BC: Pacific Educational Press.
- HOXBY, C. E. 2002. "The Cost of Accountability". Trabalho para Discussão 8855, National Board of Economic Research, Cambridge, MA.
- HUSÉN, T. 1973. "Foreword". In *Science Achievement in Nineteen Countries*, org. L. C. Comber e J. P. Keeves, 13–24. New York: Wiley.
- HUSÉN, T. e T. N. Postlethwaite. 1996. "A Brief History of the International Association for the Evaluation of Educational Achievement (IEA)". *Assessment in Education* 3 (2): 129–41.
- IEA. 2000. *Framework and Specifications for PIRLS Assessment 2001*. Chestnut Hill, MA: International Study Center, Boston College.
- IIEP. 2007. "Southern and Eastern Africa Consortium for Monitoring Educational Quality." IIEP, Paris. <http://www.unesco.org/iiep/eng/networks/sacmeq/sacmeq.htm>.
- ILON, L. 1996. "Considerations for Costing National Assessments". In *National Assessment: Testing the System*, P. Murphy, V. Greaney, M. E. Lockheed e C. Rojas (orgs.), 69–88. Washington, DC: Banco Mundial.
- ÍNDIA, National Council of Educational Research and Training, Department of Educational Measurement and Evaluation. 2003. *Learning Achievement of Students at the End of Class V*. Nova Deli: Department of Educational Measurement and Evaluation.
- ISHINO, T. 1995. "Japan". In *Performance Standards in Education: In Search of Quality*, 149–61. Paris: OCDE.
- JOHNSON, E. G. 1992. "The Design of the National Assessment of Educational Progress". *Journal of Educational Measurement* 29 (2): 95–110.
- JONES, L. V. 2003. "National Assessment in the United States: The Evolution of a Nation's Report Card". In *International Handbook of Educational Evaluation*. T. Kellaghan e D. L. Stufflebeam (orgs.), 883–904. Dordrecht, Holanda: Kluwer Academic.
- KANJEE, A. 2006. "The State of National Assessments of Learner Achievement". Trabalho inédito preparado para o Human Sciences Research Council, Pretória, África do Sul.

- KEEVES, J. P. 1995. "The Contribution of IEA Research to Australian Education". In *Reflections on Educational Achievement: Papers in Honour of T. Neville Postlethwaite*. W. Bos e R. H. Lehmann, 137–58. Nova York: Waxman.
- KELLAGHAN, T. 1996. "IEA Studies and Educational Policy". *Assessment in Education* 3 (2): 143–60.
- . 1997. "Seguimiento de los resultados educativos nacionales". In *Evaluación y reforma educativa: Opciones de política*. B. Álvarez H. e M. Ruiz-Casares (orgs.), 23–65. Washington, DC: ABEL/PREAL/AID.
- . 2003. "Local, National and International Levels of System Evaluation: Introduction". In *International Handbook of Educational Evaluation*. T. Kellaghan e D. L. Stufflebeam, 873–82. Dordrecht, Holanda: Kluwer Academic.
- . 2006. "What Monitoring Mechanisms Can Be Used for Cross-National (and National) Studies?" In *Cross-National Studies of the Quality of Education: Planning Their Design and Managing Their Impact*. K. N. Ross e I. J. Genevois (orgs.), 51–55. Paris: International Institute for Educational Planning.
- KELLAGHAN, T. e V. Greaney. 2001a. "The Globalisation of Assessment in the 20th Century". *Assessment in Education* 8 (1): 87–102.
- . 2001b. *Using Assessment to Improve the Quality of Education*. Paris: International Institute for Educational Planning.
- . 2004. *Assessing Student Learning in Africa*. Washington, DC: Banco Mundial.
- KHANIYA, T. e J. H. Williams. 2004. "Necessary but Not Sufficient: Challenges to (Implicit) Theories of Educational Change – Reform in Nepal's Education System". *International Journal of Educational Development* 24 (3): 315–28.
- KIRSCH, I. 2001. *The International Adult Literacy Study (IALS): Understanding What Was Measured*. Princeton, NJ: Educational Testing Service.
- KULPOO, D. e P. Coustère. 1999. "Developing National Capacities for Assessment and Monitoring through Effective Partnerships". In *Partnerships for Capacity Building and Quality Improvements in Education: Papers from the ADEA 1997 Biennial Meeting, Dacar*. Paris: Associação para o Desenvolvimento da Educação na África.
- LESOTO, Examinations Council of Lesotho e National Curriculum Development Centre. 2006. *Lesoto: National Assessment of Educational Progress, 2004*. Maseru: Examinations Council of Lesotho e National Curriculum Development Centre.
- LLECE. 2002. *Qualitative Study of Schools with Outstanding Results in Seven Latin American Countries*. Santiago: LLECE.
- LOCKHEED, M. E. e A. Harris. 2005. "Beneath Education Production Functions: The Case of Primary Education in Jamaica". *Peabody Journal of Education* 80 (1): 6–28.

- MAKUWA, D. 2005. *The SACMEQ II Project in Namibia: A Study of the Conditions of Schooling and Quality of Education*. Harare: Southern and Eastern Africa Consortium for Monitoring Educational Quality.
- MCMEEKIN, R. W. 2000. *Implementing School-Based Merit Awards: Chile's Experiences*. Washington, DC: Banco Mundial.
- MICHAELOWA, K. 2001. "Primary Education Quality in Francophone Sub-Saharan Africa: Determinants of Learning Achievement and Efficiency Considerations". *World Development* 29 (10): 1699–716.
- MRUTU, A., G. Ponera e E. Nkumbi. 2005. *The SACMEQ II Project in Tanzania: A Study of the Conditions of Schooling and the Quality of Education*. Harare: Southern and Eastern Africa Consortium for Monitoring Educational Quality.
- MULLIS, I. V. S., A. M. Kennedy, M. O. Martin e M. Sainsbury. 2006. *PIRLS 2006: Assessment Framework and Specifications*. Chestnut Hill, MA: International Study Center, Boston College.
- MULLIS, I. V. S., M. O. Martin, E. J. Gonzalez e S. J. Chrostowski. 2004. *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: International Study Center, Boston College.
- MULLIS, I. V. S., M. O. Martin, E. J. Gonzalez e A. M. Kennedy. 2003. *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools*. Chestnut Hill, MA: International Study Center, Boston College.
- MULLIS, I. V. S., M. O. Martin, G. J. Ruddock, C. Y. O'Sullivan, A. Arora e E. Erberber. 2005. *TIMSS 2007 Assessment Frameworks*. Chestnut Hill, MA: International Study Center, Boston College.
- MURIMBA, S. 2005a. "The Impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)". *Prospects* 35 (1): 91–108.
- . 2005b. "The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ): Mission Approach and Projects". *Prospects* 35 (1): 75–89.
- NASSOR, S. e K. A. Mohammed. 1998. *The Quality of Education: Some Policy Suggestions Based on a Survey of Schools – Zanzibar*. SACMEQ Policy Research 4, International Institute for Educational Planning, Paris.
- NAUMANN, J. 2005. "TIMSS, PISA, PIRLS, and Low Educational Achievement in World Society". *Prospects* 35 (2): 229–48.
- OCDE. 2001. *Outcomes of Learning: Results from the 2000 Program for International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science Literacy*. Paris: OCDE. <http://nces.ed.gov/pubs/2002/2002115.pdf>.
- . 2003. *The PISA 2003 Assessment Framework: Reading, Mathematics, Science and Problem Solving Knowledge and Skills*. Paris: OCDE.

- . 2004a. *First Results from PISA 2003: Executive Summary*. Paris: OCDE. <http://www.oecd.org/dataoecd/1/63/34002454.pdf>.
- . 2004b. *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OCDE.
- . 2007. "Sample Questions: PISA Mathematics with Marking Guide". OCDE, Paris. <http://pisa-sq.acer.edu.au>.
- OCDE e Instituto de Estatísticas da UNESCO. 2003. *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*. Paris e Montreal: OCDE e Instituto de Estatísticas da UNESCO.
- OLIVARES, J. 1996. "Sistema de Medición de la Calidad de la Educación de Chile: SIMCE, Algunos Problemas de la Medición". *Revista Iberoamericana de Educación* 10. <http://www.rieoei.org/oeivirt/rie10a07.htm>.
- PASSOS, A., T. Nahara, F. Magaia e C. Lauchande. 2005. *The SACMEQ II Project in Mozambique: A Study of the Conditions of Schooling and the Quality of Educational Quality*. Harare: Southern and Eastern Africa Consortium for Monitoring Educational Quality.
- PERERA, L., S. Wijetunge, W. A. de Silva e A. A. Navaratne. 2004. *Achievement after Four Years of Schooling. National Assessment of Achievement of Grade Four Pupils in Sri Lanka: National Report*. Colombo: National Education Research and Evaluation Centre, Universidade de Colombo.
- POSTLETHWAITE, T. N. 2004. "What Do International Assessment Studies Tell Us about the Quality of School Systems?" Documento básico para o *Education for All Global Monitoring Report 2005*, UNESCO, Paris.
- PRAKASH, V., S. K. S. Gautam e I. K. Bansal. 2000. *Student Achievement under MAS: Appraisal in Phase-II States*. Nova Deli: National Council of Educational Research and Training.
- RAMIREZ, F. O., X. Luo, E. Schofer e J. W. Meyer. 2006. "Student Achievement and National Economic Growth". *American Journal of Education* 113 (1): 1–29.
- RAVELA, P. 2005. "A Formative Approach to National Assessments: The Case of Uruguay". *Prospects* 35 (1): 21–43.
- REDDY, V. 2005. "Cross-National Achievement Studies: Learning from South Africa's Participation in the Trends in International Mathematics and Science Study". *Compare* 35 (1): 63–77.
- . 2006. *Mathematics and Science Achievement at South African Schools in TIMSS 2003*. Cidade do Cabo, África do Sul: Human Sciences Research Council Press.
- ROBITAILLE, D. F., A. E. Beaton e T. Plomp (orgs.). 2000. *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science*. Vancouver, BC: Pacific Educational Press.

- ROJAS, C. e J. M. Esquivel. 1998. “Los Sistemas de Medición del Logro Académico en Latino América”. LCSHD Paper 25, Washington, DC: Banco Mundial.
- ROSS, K. 1987. “Sample Design”. *International Journal of Educational Research* 11 (1): 57–75.
- ROSS, K. e T. N. Postlethwaite. 1991. *Indicators of the Quality of Education: A Study of Zimbabwean Primary Schools*. Harare: Ministry of Education and Culture; Paris: International Institute for Educational Planning.
- SHABALALA, J. 2005. *The SACMEQ II Project in Swaziland: A Study of the Conditions of Schooling and the Quality of Education*. Harare: Southern and Eastern Africa Consortium for Monitoring Educational Quality.
- SHUKLA, S., V. P. Garg, V. K. Jain, S. Rajput e O. P. Arora. 1994. *Attainments of Primary School Children in Various States*. Nova Deli: National Council of Educational Research and Training.
- SOFRONIOU, N. e T. Kellaghan. 2004. “The Utility of Third International Mathematics and Science Study Scales in Predicting Students’ State Examination Performance”. *Journal of Educational Measurement* 41 (4): 311–29.
- STRAUS, M. 2005. “International Comparisons of Student Achievement as Indicators for Educational Policy in Slovenia”. *Prospects* 35 (2): 187–98.
- UNEB. 2006. *The Achievements of Primary School Pupils in Uganda in English Literacy and Numeracy*. Kampala: UNEB.
- UNESCO. 1990. *Final Report of the World Congress on Education for All: Meeting Basic Learning Needs, Jomtien, Thailand*. Paris: UNESCO.
- . 2000. *The Dakar Framework for Action – Education for All: Meeting Our Collective Commitments*. Paris: UNESCO.
- . 2001. *Technical Report of the First International Comparative Study*. Santiago: Escritório Regional para a América Latina e o Caribe.
- . 2002. *EFA Global Monitoring Report 2002: Is the World on Track?* Paris: UNESCO.
- . 2004. *EFA Global Monitoring Report 2005: The Quality Imperative*. Paris: UNESCO.
- WILKINS, J. L. M., M. Zembylas e K. J. Travers. 2002. “Investigating Correlates of Mathematics and Science Literacy in the Final Year of Secondary School”. In *Secondary Analysis of the TIMSS Data*, org. D. F. Robitaille e A. E. Beaton, 291–316. Dordrecht, Holanda: Kluwer Academic.
- WILLMS, J. D. e M.-A. Somers. 2005. “Raising the Learning Bar in Latin America: Measuring Student Outcomes”. Policy Brief, Canadian Research Institute for Social Policy, Universidade de New Brunswick, Fredericton.
- WINOGRAD, P. e B. Thorstensen. 2004. “Using Large Scale Assessments to Inform the Policies and Practices That Support Student Learning”. Trabalho para

discussão apresentado na International Reading Association e World Bank's Global National Assessment Training Project, Escritório de Responsabilização na Educação, Santa Fé, NM.

WOLFF, L. 1998. "Educational Assessment in Latin-America: Current Progress and Future Challenges". Trabalho para Discussão 11, Programa de Promoción de la Reforma Educativa en America Latina y el Caribe, Partnership for Educational Revitalization in the Americas, Washington, DC.

ZHANG, Y. 2006. "Urban-Rural Literacy Gaps in Sub-Saharan Africa: The Roles of Socioeconomic Status and School Quality". *Comparative Education Review* 50 (4): 581-602.



ÍNDICE

Quadros, figuras e tabelas estão indicadas com q, f e t.

- abordagens de avaliação nacional
 - baseadas no censo e em amostra, 33-34, 84, 88
- Administração Nacional da Educação Pública (Uruguai), 97
- África
 - avaliação de professores na, 14
 - objetivos da avaliação nacional na, 19
 - PIRLS na, 127
 - população a ser avaliada na, 32
 - taxas de escolarização líquida na, 79
 - TIMSS na, 72
 - Veja também* Consórcio da África Austral e Oriental para Monitoramento da Qualidade da Educação (SACMEQ)
- África do Sul, 42, 77, 80t, 82q, 88, 89, 122, 124t, 127, 139, 145
- estudo de caso da, 99-102
- África Subsaariana, taxas de escolarização líquida na, 79
- Alemanha, 68, 76, 128t, 134f, 136f, 137f
- América do Norte, matrícula no ensino fundamental, 79
- América Latina e Caribe
 - avaliação nacional na, 8
 - avaliações como substituto de exames públicos na, 34
 - dados comparativos e, 81
 - institutos de avaliação na, 85
 - padrões de frequência à escola na, 79
 - PISA na, 129, 131
 - Veja também* Laboratório Latino-Americano de Avaliação da Qualidade da Educação (LLECE)
- amostra ou população a ser avaliada, 33-34, 55t, 84, 88
- análise de conteúdo para avaliação nacional, 36

- análises e procedimentos estatísticos
- atividades recomendadas, 62
 - custos de, 53
 - decisões envolvidas na avaliação nacional e, 48, 50, 55t
 - erros comuns, 62-63
 - questões na, 61-63
- aplicação de testes/provas e questionários, 30, 53, 55t, 81
- apresentação do aproveitamento de alunos
- desempenho em domínios do currículo, 45-47
 - elaboração de relatório, 64-65
 - informações no nível de item, 46
 - padrões de desempenho, 47
 - padrões de domínio, 47-48, 49t
- aproveitamento dos professores, avaliação do, 14, 39, 92, 95, 145
- Arábia Saudita, 80t, 122, 124t
- Argentina, 4, 25, 127, 128t, 130, 154, 155t, 156t, 157
- Armênia, 124t
- Associação Internacional para Avaliação do Aproveitamento Escolar (IEA), 36-37, 69, 81, 90, 141
- Veja também* Estudo Internacional sobre o Progresso do Letramento em Leitura (PIRLS); Tendências Internacionais no Estudo de Matemática e Ciências (TIMSS)
- Austrália, 21, 74, 124t, 133, 134f, 136f, 137f
- Áustria, 133, 134f, 136f, 137f
- avaliação de resultados afetivos, 39
- avaliação internacional do aproveitamento de alunos, 3, 67-82, 119-137
- crescimento da prática de, 69-72
 - problemas com, 77-82
 - vantagens da, 72-76
- Avaliação Nacional de Leitura em Inglês (Irlanda), 42
- Avaliação Nacional do Progresso Educacional (NAEP, EUA), 19, 20f, 42, 54, 68, 110-112
- avaliação subnacional, 4
- Azerbaijão, 130
- Bahrain, 80t, 124t
- Banco Interamericano, 97
- Banco Mundial, 97
- Bangladesh, 14
- Bélgica, 124t, 134f, 136f, 137f
- Belize, 127, 128t
- Benin, 151
- Bolívia, 154, 155t, 156t, 157f
- Botsuana, 80t, 100, 124t, 139
- Brasil, 4, 33, 80, 130, 133, 134f, 136f, 137f, 154, 155t, 156t, 157
- Bulgária, 124t, 129, 128t
- Burkina Faso, 149, 151
- Butão, 17-18, 89
- Camarões, 14, 149, 151, 151f
- Cambódia, 29, 54
- Canadá, 75, 124t, 128t, 133, 134f, 136f, 137
- capacidade técnica
- compromisso de longo prazo do governo e, 85
 - condições para a avaliação nacional e, 84
 - de usuários, 86
 - divulgação e uso de avaliação nacional e, 50
 - doadores e, 54
 - efeito da falta de, 39
 - estudos internacionais e, 90
 - procedimentos estatísticos e, 61-63, 88
 - realização de avaliação nacional e, 25, 26t, 27t
- Caribe, países do. *Veja* América Latina e Caribe
- Cazaquistão, 14

- Chade, 14, 151
- Chile
- abordagens de avaliação baseadas no censo no, 34
 - apresentação e uso de resultados no, 50
 - avaliação de resultados afetivos no, 39
 - compromisso do governo com o processo de avaliação no, 85
 - estudo de caso do, 88, 89, 107-110
 - frequência da avaliação no, 45
 - LLECE no, 154, 155t, 156t, 157f
 - objetivos da avaliação nacional no, 21, 22
 - órgão de implementação no, 88
 - PISA no, 130, 131
 - programa de premiação no, 21, 22, 107-110
 - TIMSS no, 122
- China, 122, 123, 127, 130, 130
- Chipre, 124t, 128t
- Cingapura, 75, 80t, 122, 123, 127, 128t
- classificação de pontuações de testes, 63, 73, 80-81
- Veja também* tabela classificativa
- Colômbia, 32, 33, 39, 122, 127, 128t, 130, 154, 155t, 156t, 157f
- comissão de coordenação nacional, 23-25, 28, 33, 39, 51, 55t
- comissão nacional de exames, 88
- compromisso de longo prazo do governo, importância do, 85
- CONFEMEN (Conferência dos Ministros da Educação dos Países de Língua Francesa em Porcentagem), 148-149
- Veja também* Programa de Análise dos Sistemas Educacionais da CONFEMEN (PASEC)
- confiabilidade, 14, 45, 53, 87, 106
- Congo, República Democrática do, 14, 151
- Consórcio da África Austral e Oriental para Monitoramento da Qualidade da Educação (SACMEQ), 139-148
- avaliação do aproveitamento de professores, 39
 - constatações, 145-148
 - dados comparativos e, 81
 - instrumentos, 47, 140-145
 - marco de referência, 139-140
 - participantes, 69, 100, 145
 - população avaliada e, 32
 - uso de dados do, 74
- Coreia do Sul, 75, 80t, 122, 124t, 123, 130, 133, 134f, 136, 137
- Costa do Marfim, 149, 151
- Costa Rica, 33, 154
- Cuba, 33, 124t, 154, 155t, 156t, 157
- currículo
- desempenho em domínios do, 35-36, 45-47, 62
 - desenvolvimento e avaliação, 9, 32, 38-39, 44, 51, 54, 61, 84
 - divulgação de dados de avaliação e, 46
 - equipes e, 29
 - estreitamento do, 58
 - estudos internacionais e, 68, 69, 70t, 73, 75-78
 - exames públicos e, 14, 16t, 15
 - formuladores de políticas e, 79, 87
 - frequência de avaliações e, 45
 - TIMSS e, 125
- custos, 45, 74, 82
- da aplicação de teste, 54
 - da avaliação nacional, 52-55
 - de análises e procedimentos estatísticos, 53

- necessidades de pessoal e, 53
- órgãos de implementação e, 52
- TIMSS e, 82
- decisões em avaliações nacionais, 2, 23-55
 - amostra ou censo, 33-34, 55t, 84, 88
 - aplicação de provas e questionários, 30, 53, 55t
 - comitê de coordenação nacional (CCN), 23-25, 28, 33, 39, 51, 55t
 - como o aproveitamento será avaliado, 39-44, 53, 55t
 - componentes de custo, 52-55
 - divulgação e uso de resultados, 50-51, 55t
 - frequência da avaliação, 45, 55t, 88
 - marco de referência da avaliação, 35-39, 55t
 - o que será avaliado, 35-39, 55t
 - orientações de políticas para, 23-25, 55t, 84-85
 - população a ser avaliada, 31-34, 55t, 84
 - principal responsabilidade por, 55
 - procedimentos estatísticos, 48, 50, 53, 55t
 - quem deve realizar a avaliação nacional, 2, 28-29, 55t
 - relatórios sobre aproveitamento de alunos, 45-48, 55t
 - resumo de, 47
- Dinamarca, 134f, 136f, 137f
- Direção Nacional de Informação e Avaliação da Qualidade da Educação (DiNIECE, Argentina), 28, 27
- diretrizes de políticas para a avaliação nacional, 23-28, 55t, 84-85, 89
 - Veja também* Apêndices A e C
- divulgação e uso das constatações, 2, 3, 85-86
- Djibuti, 151
- Educação Para Todos, 1, 9, 148
- Egito, 100, 124t
- elementos de uma avaliação nacional, 13-14
- Emirados Árabes Unidos, 14
- equipes para realizar a avaliação nacional, 29
- erros comuns, 57-65
- Escócia, 76, 124t, 128t
- Eslovênia, 124t, 128t
- Espanha, 75, 134f, 136f, 137f
- Estados Unidos
 - avaliação subnacional nos, 4
 - avaliações como substituto de exames públicos nos, 34
 - estudo de caso dos, 110-112
 - objetivos da avaliação nacional nos, 18-20, 110-111
 - PIRLS, 128t
 - PISA, 133, 134f, 136, 137f
 - população a ser avaliada nos, 31, 32
 - TIMSS, 75, 124t, 125
 - Veja* Avaliação Nacional do Progresso Educacional (NAEP)
- Estônia, 124t
- Estudo Internacional sobre o Progresso do Letramento em Leitura (PIRLS), 125-128
 - constatações, 127-128
 - frequência, 45
 - instrumentos, 36-38, 69, 126
 - marco de referência, 125-126
 - participantes, 127
- estudos de caso de países, 4, 5, 88-89, 91-117
 - África do Sul, 99-102
 - Chile, 107-110
 - Estados Unidos, 110-112
 - Índia, 91-93

- Nepal, 105-107
 Sri Lanka, 102-105
 Uganda, 112-117
 Uruguai, 97-99
 Vietnã, 93-96
 estudos internacionais, 3, 4, 89-90, 119-137
 Estudo Internacional sobre o Progresso do Letramento em Leitura (PIRLS), 125-128
 Programa Internacional de Avaliação de Alunos (PISA), 38-39, 42, 129-137
 Tendências Internacionais no Estudo de Matemática e Ciências (TIMSS), 34, 42, 90, 119-123
 estudos regionais, 3, 4, 90, 139-158
 Consórcio da África Austral e Oriental para Monitoramento da Qualidade da Educação (SACMEQ), 139-148
 Laboratório Latino-Americano de Avaliação da Qualidade da Educação (LLECE), 153-157
 Programa de Análise dos Sistemas Educacionais da CONFEMEN (PASEC), 148-152
 Etiópia, 11, 29, 50
 Europa Ocidental, taxas de escolarização líquida na, 79
 exames públicos *vs.* avaliação nacional do aproveitamento de alunos, 7-8, 14-15
 fatores contextuais e aproveitamento de alunos, 39-39
 Federação Russa, 68, 75, 124t, 128t, 134f, 136f, 137f
 Filipinas, 75, 80t, 122, 124t
 Finlândia, 133, 134f, 136-137
 Formato de múltipla escolha, 42, 42q
 França, 33, 128t, 134f, 136f, 137f
 frequência da avaliação, 45, 55t, 88, 91, 93, 97, 101, 105, 107, 111, 113, 119, 126, 130, 145, 149
 Gabão, 151
 Gana, 44, 46, 79, 80t, 100, 124t
 gênero
 casos de países e, 91, 92, 96t, 104, 106, 113, 115
 educação em matemática e ciências e, 21
 frequência à escola e, 79
 PASEC e, 153
 PIRLS e, 129
 PISA e, 133, 136
 SACMEQ e, 142, 147, 148, 151
 TIMSS e, 125
 uso da avaliação nacional para apoiar ações de políticas e, 89
 globalização e avaliação internacional, 1, 72
 Grécia, 128t, 136f, 137f
 Guiné, 14, 79, 149, 151
 habilidades para a vida, avaliação de, 39
 Honduras, 33, 154, 155t, 156t, 157f
 Hong Kong, 122, 124t, 123, 127, 128t, 130, 136f, 137f
 Hungria, 124t, 128t, 134f, 136f, 137f
 idioma (língua) da avaliação, 9, 33, 42-44, 48, 78, 101, 102, 104, 109t, 115, 129
 IEA. *Veja* Associação Internacional para Avaliação do Aproveitamento Escolar
 Iêmen, 79, 122
 implementação, 3, 57-65
 opções para a, 28, 26-27t
 Índia, 14, 79, 89
 estudo de caso da, 91-93
 Indonésia, 80, 122, 124t, 127, 130, 133, 136f, 137
 informações no nível de item, divulgação de, 39, 45-46, 92, 98

- Inglaterra, 45, 75, 124t, 128t
 institucionalização da avaliação nacional, 85
 Instituto Internacional para o Planejamento Educacional (IIEP), 139
Veja também Consórcio da África Austral e Oriental para Monitoramento da Qualidade da Educação (SACMEQ)
 institutos de pesquisa, 25, 26t, 29, 88
 Irã, 122, 124t, 127, 128t
 Irlanda, 42, 76, 134f, 136f, 137f
 Islândia, 76, 128t, 134f, 136f, 137f
 Israel, 122, 124t, 127, 128t, 130
 Itália, 124t, 128t, 134f, 136f, 137f
 Jamaica, 33
 Japão, 45, 75, 122, 124t, 130, 133, 134f, 136, 137f
 Jomtien, Declaração de (1990), 9
 Jordânia, 33, 130
 Kuwait, 75, 122, 127, 128t
 Laboratório Latino-Americano de Avaliação da Qualidade da Educação (LLECE)
 constatações, 153-158
 instrumentos, 153
 marco de referência, 153
 participantes, 154
 Laos, 14
 Lesoto, 47, 51, 139
 Letônia, 76, 124t, 123, 128t, 134f, 136f, 137f
 letramento, marco de avaliação do, 36-38
Veja também Estudo Internacional sobre o Progresso do Letramento em Leitura (PIRLS)
 Líbano, 124t
 Liechtenstein, 134f, 136f, 137f
 Lituânia, 124t, 123, 128t
 Luxemburgo, 134f, 136f, 137f
 Macau, 130, 136f, 137f
 Macedônia, 124t, 128t
 Madagascar, 151
 Malásia, 122, 124t
 Malawi, 19, 139, 145, 147
 Maldivas, 30
 Mali, 149, 151
 Marco de Referência e Especificações (IEA), 36-37
 marcos de referência para a avaliação, 35-39, 55t
 Marrocos, 80t, 100, 122, 124t, 127, 128t
 matrícula no ensino fundamental, 79
 Maurício, 139, 145, 146, 147f, 148, 151
 Mauritânia, 14, 151
 MESyFOD (Programa de Modernização da Educação Secundária e Formação de Docentes), 97
 México, 33, 34, 130, 134f, 136f, 137f, 154, 155t, 156t, 157f
 mídia e avaliações internacionais, 73
 Ministérios da Educação
 como órgãos implementadores, 88
 custos da avaliação nacional e, 54, 55
 divulgação e uso de resultados e, 51
 o que será avaliado, definição de, 39
 realização de avaliação nacional pelos, 2, 25, 26t, 29
 Moçambique, 79, 139, 145
 Moldávia, 124t, 128t
 Monitoramento do Aproveitamento do Aprendizado, 39, 100, 104
 Namíbia, 19, 139, 146, 147
 Nepal, 19, 88
 estudo de caso do, 105-107
 Níger, 149, 151

- Nigéria, 33
 Noruega, 76, 124t, 128t, 133, 134f, 136f, 137f
 Nova Zelândia, 75, 124t, 128t, 134f, 136f, 137f
 objetivos da avaliação nacional, 2, 17-22
 Objetivos de Desenvolvimento do Milênio, 22
 Organização para a Cooperação e o Desenvolvimento Econômico (OCDE), 69, 129
Veja também Programa Internacional de Avaliação de Alunos (PISA)
 órgãos de implementação
 aplicação de testes e questionários por, 31
 custos e, 52
 decisões, resumo das, 55t
 divulgação e uso de resultados e, 51
 o que será avaliado, determinação de, 34
 realização de avaliação nacional e, 25, 26-27t, 29
 responsabilidade por decidir como o aproveitamento será avaliado, 45
 padrão de referência, 72, 79, 80t, 124t
 padrões de domínio, 47-48, 49t
 País Basco, 124t
 países árabes, padrões de frequência à escola nos, 79
 Países Baixos, 124t, 129, 128t, 133, 136f, 137f
 Palestina, 124t
 Paquistão, 14
 Paraguai, 154, 155t, 156t, 157f
 PASEC. *Veja* Programa de Análise dos Sistemas Educacionais da CONFEMEN
 Peru, 14, 130, 130, 154, 155t, 156t
 PIRLS. *Veja* Estudo Internacional sobre o Progresso do Letramento em Leitura
 PISA. *Veja* Programa Internacional de Avaliação de Alunos
 Plano de Ação de Dacar (2000), 9
 Polônia, 133, 134f, 136f, 137f
 população a ser avaliada
 baseada no censo ou em amostra, 33-34, 55t, 84, 88
 decisões na avaliação nacional e, 31-33, 55t
 na avaliação internacional, 78-79
 subgrupos, 2, 9
 Portugal, 136f, 137f
 pré-teste, 4, 28
 Veja também teste-piloto
 Produtos da educação, ênfase nos, 1-2
 Programa de Análise dos Sistemas Educacionais da CONFEMEN (PASEC), 148-152
 constatações, 151-152
 instrumentos, 150
 marco de referência, 148-149
 participantes, 151
 população a ser avaliada e, 32
 Programa de Modernização da Educação Secundária e Formação de Docentes (MESyFOD), 97
 Programa Internacional de Avaliação de Alunos (PISA), 129-137
 comparação com o TIMSS, 69, 70-71t
 constatações, 72, 90, 133-137
 instrumentos, 42, 131-133
 marco de referência da avaliação do, 38-39
 marco de referência, 90, 129
 participantes, 69, 97, 130-131
 uso de dados do, 68, 74, 75-76

- programas de prêmios por desempenho, 21, 22, 107-110
- projeto de uma avaliação nacional, 2-4, 57-65
- Qatar, 127, 130
- Quênia, 48, 139, 146, 147f, 148
- questões abertas na avaliação, 42, 43q
- Quirguistão, 130
- realização de avaliação nacional, 2, 28-29, 55t
- Reino Unido, 134f
- República Centro-Africana, 151
- República Dominicana, 14, 154, 155t, 156t, 157f
- República Eslovaca, 76, 124t, 128t, 136f, 137f
- República Tcheca, 75, 128t, 134f, 136f, 137f
- responsabilidade pela avaliação nacional, 26-27t, 55t
Veja também Apêndice A
- România, 124t, 128t
- SACMEQ.
Veja Consórcio da África Austral e Oriental para Monitoramento da Qualidade da Educação
- sanções e desempenho, 22, 33-34
- Seicheles, 139, 145, 146
- Senegal, 149, 151
- Serra Leoa, 24
- Sérvia, 124t, 136f, 137
- Síria, 122
- Sistema de Medição da Qualidade da Educação (SIMCE), 85, 107-110
- situação econômica e aproveitamento dos alunos, 39, 63, 72, 76, 78, 90, 98, 103t, 108, 146, 157
- SNED (Sistema Nacional de Avaliação do Desempenho dos Estabelecimentos Educacionais Subvencionados, Chile), 107-110
- Sri Lanka, 32, 48, 51, 88
estudo de caso de, 102-105
- Suazilândia, 139, 145
- Suécia, 75, 124t, 129, 128t, 134f, 136f, 137f
- Suíça, 75, 76, 134f, 136f, 137f
- tabelas classificativas, 22, 35t, 73, 133
Veja também classificação de pontuações de testes
- Tailândia, 122, 130, 136f, 137f
- Taipei, 80t, 122, 124t, 127, 130
- Tanzânia, 139, 145, 146, 147
- Tendências Internacionais no Estudo de Matemática e Ciências (TIMSS), 119-123
comparado com o PISA, 69, 70-71t
constatações, 90, 122-123
custo do, 82
marco de referência, 90, 119-120
motivação dos estudantes e, 34
participantes, 69, 72, 99-101, 122
SACMEQ e, 140-141
uso de dados do, 68, 74-75
- teste-piloto, 5, 13, 28, 54, 59, 61, 121
Veja também pré-teste
- testes e questionários, aplicação de, 30, 53, 55t, 81
- Togo, 149, 151
- treinamento de professores, 1, 6, 7, 14, 18, 20, 24, 38, 51, 64, 73, 74, 104, 106, 111, 123, 149, 151, 152
durante o trabalho, 53, 98, 152
- Trinidad e Tobago, 127
- Tunísia, 80, 100, 122, 124t, 130, 133, 136f, 137

- Turquia, 127, 128t, 130, 136f, 137f
- Uganda, 39, 42, 88, 139
 estudo de caso de, 113-117
- UNESCO, 139
- Unidade de Medição de Resultados Educacionais (UMRE), 97-99
- universidades e avaliação nacional, 8, 12q, 25, 26t, 29, 74
- Uruguai
 censo v. amostra, abordagens da avaliação nacional no, 32, 33
 contribuição para a formulação de políticas no, 88
 divulgação e uso de resultados no, 51
 estudo de caso do, 88, 88, 97-99
 objetivos da avaliação nacional no, 21, 22
 PISA, 131, 136f, 137f
- validade, 4, 5, 13, 34, 44, 51, 61, 64, 76, 86
- Venezuela, 154, 155t, 156t, 157f
- Vietnã
 aproveitamento de professores no, 14, 39, 95
 contribuição para a formulação de políticas no, 88
 divulgação e uso de resultados no, 51
 estudo de caso do, 88, 93-96
 exemplos de questões na avaliação nacional do, 10, 11q
 objetivos da avaliação nacional no, 17
 padrões de desempenho no, 47
 procedimentos estatísticos no, 48, 88
- Zâmbia, 14, 19, 30, 139, 146, 147, 148
- Zanzibar, 17, 139, 147
- Zimbábue, 139, 140, 141, 145, 146f, 147, 148

AVALIAÇÃO DOS NÍVEIS DE DESEMPENHO EDUCACIONAL

descreve os objetivos e os vários aspectos das avaliações nacionais e examina questões relativas a projeto, implementação, análise, elaboração de relatórios e disseminação de resultados. Também apresenta importantes avaliações internacionais, regionais e nacionais realizadas em diversos países.

Este livro interessa a órgãos governamentais nacionais, estaduais e municipais, além de instituições de pesquisa e universidades.

