

PREAL

Programa de Promoção da Reforma Educativa na América Latina e Caribe

Partnership for Educational Revitalization in the Americas

Nº 11

**Avaliações educacionais
na América Latina:
estágio atual e desafios
futuros**

*Laurence Wolff**

Os documentos desta série estão disponíveis em formato eletrônico na Internet (www.preal.cl).

Traduzido por Paulo Martins Garchet - (E-mail: garchet@unisys.com.br)

Revisão Técnica: Helena Maria B. Bomeny

* Laurence Wolff é consultor para educação do Banco Interamericano de Desenvolvimento, em Washington, D.C. Trabalhou anteriormente no Banco Mundial durante 22 anos, período em que esteve envolvido em projetos de educação na América Latina, na Ásia e no Oriente Médio. Tem doutorado em educação pela Universidade de Harvard e escreveu o presente relatório enquanto esteve como visitante no Education Development Center, um instituto sem fins lucrativos para pesquisa e desenvolvimento de educação com sede em Newton, Massachusetts.

O apoio para este projeto foi dado pelo Diálogo Interamericano através do Programa para Reforma Educacional na América Latina e Caribe. O PREAL é co-dirigido pelo Diálogo Interamericano em Washington, D.C. e pela Corporação Para Desenvolvimento de Pesquisa - CINDE, Santiago de Chile. É financiado pela Agência dos Estados Unidos Para Desenvolvimento Econômico (USAID - U.S. Agency for International *Development*), pelo Centro Internacional de Desenvolvimento de Pesquisas do Canadá (IDRC - *International Development Research Centre*), pelo Banco Interamericano de Desenvolvimento (BID), pelo Fundo GE (*GE Fund*) e outros doadores.

Índice

I. Introdução

II. Recentes Progressos em Avaliações Nacionais e Internacionais

Programas Internacionais

Experiências nos Estados Unidos

A Experiência Francesa

Os Países em Desenvolvimento

III. A Experiência Latino-americana

Chile

Costa Rica

Colômbia

Brasil

México

Argentina

O Programa de Avaliação Regional das UNESCO/Orealc

IV. As Lições Aprendidas e os Desafios Futuros

Resumo da Experiência em Seis Países

Construção de Consenso e Compromisso

Usos das Avaliações

Desenvolvimento de Capacidade e Competência Técnica

Novas Abordagens

Bibliografia

Apêndice: Breve Análise dos Sistemas de Avaliação em Seis Países

I. INTRODUÇÃO

A educação é fundamental para os desenvolvimentos econômico, social e cultural, para não mencionarmos a estabilidade política, a identidade nacional e a coesão social. Mais ainda, os negócios altamente tecnológicos de nossos dias não podem prosperar sem pessoas que possam contribuir com competência analítica, criatividade e cooperação para o trabalho. A existência de pessoal com estas qualificações pode ter grande impacto, ainda, na capacidade de um país atrair investimentos estrangeiros. O desempenho econômico do leste asiático, para tomarmos apenas um exemplo, pode ser em grande parte atribuído ao nível e à qualidade superiores da educação em toda esta região, que hoje se orgulha de ter quatro dos cinco melhores recordes mundiais em matemática da oitava série.

Nos últimos anos, países de todo o mundo chegaram a um acordo sobre a importância de se aferir o desempenho educacional. Acreditam que através da avaliação dos níveis correntes de desempenho e da identificação dos obstáculos ao

progresso poderão aprimorar o tipo, a profundidade e a abrangência da educação que oferecem. As avaliações educacionais podem ser definidas como aferições da medida em que foram atingidas as metas curriculares estabelecidas, seja por autoridades governamentais, seja por especialistas, nacionais ou internacionais. As avaliações nacionais aferem o progresso de instituições em todo o país e são substancialmente diferentes dos exames de conclusão ou admissão desenhados para selecionar os estudantes que serão admitidos em outro nível de educação. As avaliações internacionais comparam os resultados de aprendizagem entre países. É claro que a aferição do grau de aprendizagem dos estudantes não acarretará, por si só, um melhor desempenho dos alunos, assim como não se aumenta a produção agrícola pelo simples fato de pesarem-se os grãos colhidos. Mas é uma condição necessária para que se possam estabelecer metas quantitativas, avaliar as perdas que estratégias alternativas de alocação de recursos e combinação de insumos implicam e para que se possam orientar recursos e esforços para atingir as metas estabelecidas. Para garantir que as

avaliações educacionais, além de melhorar o aprendizado dos estudantes, sejam também eficientes em termos de custos, são precisos compromisso, financiamento adequado, conhecimento técnico, *know-how* gerencial, e agudeza política.

As avaliações educacionais são apenas um dos meios de monitorar o avanço em direção às metas educacionais. É importante avaliar também os resultados quantitativos: o número de alunos matriculados, ou que concluem determinado nível, ou que são promovidos, por exemplo; a adequação dos insumos como os livros-textos, professores, as proporções professor-aluno e o treinamento dos professores; as interações nas salas de aulas e a pedagogia; e o desempenho no mercado de trabalho (quantos graduados conseguem emprego, e a que salários, por exemplo).

As avaliações educacionais não são novidade para a América Latina. Pelo final dos anos setenta, o *Programa de Estudios Conjuntos para la Integración Económica Latinoamericana* – ECIEL, realizou um estudo comparativo do aprendizado

em cinco países, utilizando um instrumento desenvolvido pela *International Association for the Evaluation of Student Achievement* – IEA. Vários países latino-americanos participaram de avaliações internacionais promovidas, tanto pela IEA, quanto pelo *Educational Testing Service* – ETS. Os resultados de tais avaliações foram desapontadores: os pontos dos países da América Latina estiveram, consistentemente, muito abaixo dos da América do Norte, da Europa e da Ásia. No estudo da IEA de 1989 (Tabela 1), a Venezuela pontuou abaixo de todos os demais participantes, inclusive da Indonésia. Trinidad-Tobago esteve melhor que a Venezuela, mas ainda muito abaixo dos países desenvolvidos. Em um estudo de matemática em 1991 (Tabela 2), implementado pelo ETS, a pontuação das cidades de Fortaleza e São Paulo, no Brasil, ficou abaixo da de todos os países e cidades participantes, com exceção de Moçambique. Como se observa mais adiante neste relatório, a Colômbia – único país latino-americano participante – teve também um pobre desempenho no mais recente estudo de matemática da IEA.

Por volta de 1991, Costa Rica, México, Chile e Colômbia tinham

sistemas de avaliação implantados. Desde então todos os países da América Latina iniciaram algum tipo de programa. Com doação do Banco Interamericano de Desenvolvimento (BID), a UNESCO apoiou um programa regional de avaliação em

leitura e matemática de estudantes das 3^a e 4^a séries. Além disso, as avaliações internacionais cresceram, em profundidade e no número de países participantes.

TABELA 1
Estudo da IEA sobre capacidade de leitura, 1989

<u>Países da América Latina e Caribe</u>	Média	Desvio Padrão
Trinidad-Tobago	451	79
Venezuela	383	74
<u>Outros Países</u>		
Finlândia	569	70
Estado Unidos	547	74
Hong Kong	517	71
Cingapura	515	72
Espanha	504	78
Alemanha Ocidental	503	84
Indonésia	394	59

Fonte: W.B. Elley, *How in the World do Students Read?* (Newark, Delaware International Reading Association, 1992)

Tabela 2		
O Teste de Avaliação Internacional de Progresso Educacional em Matemática e Ciências, 1991		
<i>Teste de Matemática</i>		
<u>País</u>	<u>Média</u>	<u>Desvio Padrão</u>
Coréia	73	0.6
Formosa	73	0.7
Espanha	55	0.8
Estados Unidos	55	1.0
Portugal	48	0.8
São Paulo, Brasil	37	0.8
Fortaleza, Brasil	32	0.6
Maputo e Beira, Moçambique	28	0.3
<i>Teste de Ciências</i>		
<u>País</u>	<u>Média</u>	<u>Desvio Padrão</u>
Coréia	78	0.5
Formosa	76	0.4
Espanha	68	0.6
Estados Unidos	67	1.0
Portugal	63	0.8
São Paulo, Brasil	53	0.6
Fortaleza, Brasil	46	0.6
Fonte: <i>Educational Testing Service</i> , The International Assessment of Educational Progress: Learning Mathematics (Princeton, NJ. ETS, 1992)		

Este relatório analisa os progressos das avaliações educacionais, no nível internacional e em seis países da América Latina; apresenta ainda comentários sobre algumas novas direções possíveis para políticas e programas de avaliação na região. A discussão tira subsídios de recentes estudos internacionais e nacionais (identificados na bibliografia), diversos deles preparados para o seminário

realizado pelo PREAL no Rio de Janeiro, em dezembro de 1996.

II PROGRESSOS RECENTES DAS AVALIAÇÕES EDUCACIONAIS, INTERNACIONAIS E NACIONAIS

Programas Internacionais

O mais importante evento entre as avaliações internacionais em anos recentes foi o TIMMS, *Third International Mathematics and Science*

Surveys (Terceira Pesquisa Internacional sobre Matemática e Ciências) da IEA. A IEA é um consórcio para avaliações em todo o mundo, com sede em Amsterdã, na Holanda. É conhecida por seus programas de estudos internacionais sobre matemática, ciências, habilidade de leitura, alfabetização e estudos sociais. Os programas da IEA são financiados pelos países participantes.

O TIMSS procura aferir, comparar e explicar o aprendizado de ciências e matemática em 41 países. Os resultados dos exames de ciências e matemática das crianças das 4^a, 7^a, 8^a e 12^a séries estão agora disponíveis. Além disso, o TIMSS concluiu recentemente uma inovadora análise das “oportunidades de aprendizado” das crianças, que categoriza e compara o currículo, os livros-textos e a pedagogia de salas de aulas entre os 41 países participantes. Tal análise demonstra que os exames de resultados não precisam se concentrar apenas na identificação dos países com pontuação alta, ou baixa, mas podem ser usados como instrumento para aferir o progresso educacional de um

país, redefinir suas metas curriculares e mudar suas práticas de sala de aulas.

Em 1991, onze países da América Latina (Costa Rica, Peru, Argentina, República Dominicana, Colômbia, Guatemala, Venezuela, Chile, Equador, Brasil e México) compareceram a uma reunião regional preliminar do TIMSS, mas apenas Colômbia e México participaram até o fim. Além disso, apenas a pontuação da Colômbia foi apresentada na publicação do TIMSS: no último minuto, o México decidiu não permitir a publicação de seus resultados. A Argentina e a República Dominicana participaram na análise do currículo. A razão para a baixa participação dos países latino-americanos (e de todos os países em desenvolvimento) parece ser uma combinação de recursos técnicos e financeiros inadequados com, talvez, um desejo equivocado de algumas autoridades, de não arriscarem apresentar a pontuação mais baixa.

Os pontos do TIMSS referentes ao progresso em matemática na 8^a série são apresentados na Tabela 3, a seguir:

Tabela 3: Progresso em Matemática no Terceiro Estudo Internacional de Matemática e Ciências
Percentual de Estudantes que Alcançaram Níveis Internacionais de Notas, 8ª Série

País	Progresso em Matemática 8ª Série	Nível dos 10% mais altos	Nível dos 50% mais altos	Diferença de Progresso entre 7ª e 8ª Séries
Cingapura	643	45	94	42
Coreia	607	34	82	30
Japão	605	32	83	34
Hong Kong	588	27	80	24
Bélgica (Fl)	565	17	73	8
República Checa	564	18	70	40
República Eslovaca	547	12	64	39
Suíça	545	11	65	40
Áustria	539	11	61	30
França	538	7	63	46
Hungria	537	11	60	35
Eslovênia	541	11	61	43
Federação Russa	541	11	61	43
Holanda	541	10	63	25
Bulgária	540	6	58	26
Canadá	527	7	58	33
Irlanda	527	9	57	28
Bélgica (Fr)	526	6	58	19
Áustria	530	11	57	32
Tailândia	522	7	54	28
Israel	522	6	56	-
Suécia	519	5	53	41
Alemanha	509	6	49	25
Nova Zelândia	508	6	48	36
Inglaterra	506	7	48	30
Noruega	503	4	46	43
Estados Unidos	500	5	45	24
Dinamarca	502	4	47	37
Escócia	498	5	44	36
Letônia (LSS)	493	3	40	32
Espanha	487	2	36	39
Islândia	487	1	37	27
Grécia	484	3	37	44
România	482	3	36	27
Lituânia	477	1	34	49
Chipre	474	0	19	31
Portugal	454	0	19	31
Rep. Isl. do Irã	428	0	9	27
Kuwait	392	0	4	16
COLÔMBIA	385	0	4	16
África do Sul	384	0	2	7

Fonte: *Mathematics Achievements in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)* (Chestnut Hill, Mass.: Boston College/TIMSS International Study Center, 1996).

Entre os 41 países que informaram seus resultados, a Colômbia aparece em penúltimo lugar, logo acima da África do Sul. Os dados indicam que apenas 4% de todos os estudantes colombianos tiveram pontuação entre as 50% mais altas do

mundo; e nenhum dos estudantes da amostra colombiana pontuou entre os 10% mais altos do mundo, o que é digno de nota. A grande maioria dos participantes eram países desenvolvidos, com rendas per capita muito superiores às da Colômbia e a

maioria deles gastava significativamente mais em seu sistema de educação. Mesmo países com renda per capita similares – notadamente Bulgária, Lituânia, Filipinas, Romênia, Letônia, Irã, Eslováquia, Rússia e Tailândia – tiveram pontos mais altos que a Colômbia.

A metodologia do TIMSS permite aos encarregados de formular políticas e elaborar currículos, verificar se o currículo, os livros-textos e o ensino e aprendizado em salas de

aulas estão alinhados. Em consequência, os detalhes do comportamento da Colômbia em uma variedade de questões de matemática, conforme indicado na Tabela 4, dá uma boa idéia dos pontos fortes e fracos de seu currículo de matemática. Assim, os colombianos podem usar seu desempenho para comparar a qualidade de seus objetivos educacionais, currículos e livros-textos com os de outras nações, identificando as áreas que precisam melhoramentos.

Tabela 4	
Como varia o desempenho da Colômbia em relação aos currículos, pretendido e real?	
<u>Conteúdo</u>	<u>Classificação Internacional</u>
Perímetro, área e volume	1
Padrões, relações e funções	41
Geometria: transformações	41
Números reais, suas propriedades e subcategorias*	12
Números racionais e suas propriedades*	12
Frações e decimais	40
Equações e fórmulas	40
*Enfatizado nos livros-textos e pelos professores colombianos	
Fonte: "Variations in Mathematics Performance in Colombia in Accordance with the Intended and Implemented Curriculum", apresentação feita por Gilbert Valverde, "Las políticas educacionales nacionales y la participación en estudios internacionales", seminário do PREAL no Rio de Janeiro, dezembro de 1996.	

Embora o principal programa do TIMSS esteja em vias de ser concluído, uma série de atividades de *follow-up* já está em andamento. Os pesquisadores do TIMSS começaram a balizar os currículos e testes existentes (*benchmarking*) com base no exame e nas análises do TIMSS.

Além disso, há já um acordo para que o TIMSS seja repetido em 1999.

Experiência nos Estados Unidos

Os Estados Unidos tomaram a dianteira no desenvolvimento de uma ampla variedade de avaliações educacionais para aferir o

desempenho institucional, estabelecer níveis mínimos de competência para conclusão escolar e para estabelecer padrões para medir o desempenho educacional em comparação com padrões internacionais. A extensão, a complexidade e talvez os excessos dos programas de avaliação dos Estados Unidos podem ser vistos no Quadro 1, onde se vêem informações sobre quatro programas de avaliação no condado de Montgomery, Maryland. O primeiro foi organizado em torno de um teste desenvolvido pelo condado para fornecer aos pais informações sobre o nível de aprendizado nos elementos-chaves do currículo atingido por seus filhos, em suas respectivas séries, entre a 3ª e a 8ª. Este programa serve como instrumento de diagnóstico na identificação das forças e fraquezas de cada aluno e utiliza perguntas de múltipla escolha, abertas e baseadas no desempenho. O segundo programa, o *Maryland School Performance Assessment Program* (Programa de Avaliação Escolar de Maryland), afere, nas 3ª, 5ª e 8ª séries, até que ponto as escolas individuais estão atingindo os padrões de desempenho estabelecidos pelo estado. Suas perguntas são, principalmente, do tipo múltipla

escolha, ou exigem apenas respostas curtas. O terceiro programa de avaliação, o *Maryland Functional Tests* (Testes Funcionais de Maryland), é um “exame de competência mínima” em leitura, matemática, redação e cidadania porque passam obrigatoriamente todos os estudantes para receber um diploma reconhecido de secundário. O quarto é composto pelos recém-lançados *Comprehensive Tests of Basic Skills* (Testes Compreensivos de Habilidades Básicas), aplicados a uma pequena amostra dos alunos nas 3ª, 4ª e 6ª séries. Estes testes servirão como indicadores dos níveis de competências básicas em relação a normas nacionais e internacionais.

Historicamente, e por lei, os Estados Unidos não têm nenhum currículo nacional obrigatório. Como a maioria dos estados se limita a fornecer diretrizes amplas, os distritos escolares têm, individualmente, ampla margem de manobra no estabelecimento de seus currículos. De acordo com um recente relatório do U.S. National Research Center (1997), os currículos estaduais são, geralmente, parcamente definidos e pouco considerados no nível local. Como resultado, o que se ensina em sala de aula varia grandemente de

escola para escola, bem como de distrito para distrito, e alunos e professores têm, freqüentemente, uma idéia exagerada a respeito da qualidade de seus próprios desempenhos. Por isso, as autoridades se esforçam para estabelecer padrões estaduais mais fortes de aprendizado e progresso, bem como, possivelmente, padrões nacionais voluntários. De modo

similar, alguns estados – Maryland entre eles – estão tentando basear seus currículos e testes em padrões internacionais, especialmente os do TIMSS. Deve-se notar que os dois principais sindicatos de professores, especialmente a *American Federation of Teachers* (AFT), têm apoiado consistentemente os esforços estaduais e nacionais pela melhoria da qualidade e da avaliação do ensino.

Quadro 1

Avaliação Educacional nos Estados Unidos
O Programa de Avaliação do Condado de Montgomery, Maryland

O sistema de escolas públicas do Condado de Montgomery, subúrbio próximo a Washington, D.C., utiliza uma ampla bateria de testes para medir o desempenho de seus estudantes. A descrição do programa de testes que apresentamos a seguir foi retirada de um folheto informativo enviado pelo condado aos pais das crianças matriculadas em suas escolas públicas.

Avaliações do Condado de Montgomery

Testes com Referência em Critérios (CRTs). Desenvolvido pelas escolas pública do Condado de Montgomery, estes testes são aplicados todas as primaveras aos alunos da 3ª à 8ª séries para aferir até que ponto estão aprendendo e aplicando as competências ensinadas nos currículos de matemática e leitura/artes da linguagem.

Além do formato tradicional de múltipla escolha para avaliar o conhecimento de conteúdo baseado em fatos, os CRTs também incluem itens abertos e baseados no desempenho. Estes itens refletem os compromissos, nos níveis local e estadual, com instrumentos mais novos, não de múltipla escolha, que avaliam habilidades mais ricas e complexas, como as capacidades de comunicação escrita e de solução de problemas da vida real.

No outono, os pais recebem um relatório individualizado mostrando o desempenho de seus filhos, em suas respectivas séries, nos itens de múltipla escolha dos CRTs, e como seus desempenhos se comparam com os demais alunos, da escola e do condado. O relatório inclui as notas de anos anteriores, para ajudar a estimar o progresso da escola e do aluno, para identificar pontos fortes e fracos e avaliar o efeito de quaisquer medidas adotadas ao longo do ano para melhorar o desempenho.

Há planos em andamento de utilização de uma nota única que combine o desempenho do aluno nos perguntas de múltipla escolha e abertas para informar à escola e aos pais o que os alunos sabem e são capazes de fazer.

Avaliações do Estado de Maryland

Maryland School Performance Assessment Program (MSPAP). Estes testes utilizam perguntas de resposta curta e atividades ampliadas para determinar se as escolas atingiram os padrões de desempenho para leitura, redação, uso da língua, matemática, ciências e estudos sociais. Os alunos da 3ª, 5ª e 8ª séries fazem estes testes todas as primaveras. As notas individuais não são informadas neste momento. Os resultados das escolas são liberados em janeiro.

Maryland Functional Tests. Estes testes avaliam as habilidades mínimas em competências em leitura, matemática, redação e cidadania. A aprovação no teste é uma exigência para concessão do diploma do secundário em Maryland. Os estudantes das escolas públicas de Maryland fazem provas de matemática e leitura na 7ª série, redação na 8ª e cidadania na 10ª. Todos os anos, os alunos têm várias oportunidades de tentar estes testes. Informam-se as notas individuais dos alunos em todos os testes.

Comprehensive Tests of Basic Skills (CTBS). Maryland utiliza este teste como um indicador de progresso em competências básicas em leitura, língua e matemática em comparação a normas nacionais. A partir da primavera de 1997, Maryland aplicará estes testes – ano sim, ano não – nas 2ª, 4ª e 6ª séries. As escolas do Condado de Montgomery aplicarão estes testes a uma amostra de 250 estudantes em cada série. As notas individuais dos alunos serão informadas aos pais no outono.

Fonte: “Celebrating Knowledge through Assessment”, Montgomery County Public Schools (documento mimeografado), 1996.

A Experiência Francesa

A partir de meados dos anos oitenta a França desenvolveu um sofisticado sistema de avaliação estudantil. Conquanto haja um currículo nacional obrigatório, as escolas primárias e secundárias francesas têm um alto grau de liberdade na utilização de seus fundos e na seleção das técnicas pedagógicas que aplicam. A responsabilidade pelo monitoramento e pela avaliação da educação está a cargo dos mais de 200 técnicos da *Direction de l'Évaluation et de la Prospective* (DEP), que é uma divisão do Ministério da Educação francês. As mais altas autoridades governamentais estão fortemente comprometidas com a construção de um sistema de avaliação dos resultados da educação baseado em padrões claros e com a divulgação integral e aberta dos resultados, assegurando, assim, a independência da DEP.

No sistema francês, os padrões de progresso são estabelecidos centralmente e os testes são aplicados anualmente a todos os estudantes em séries específicas do primário e do secundário inferior. O

exame de conclusão (*baccalauréat*), aplicado a todos os alunos do secundário superior, é também usado para fins de avaliação. A DEP fornece *feedback* detalhado, em formato de fácil leitura, a todas as partes interessadas: público em geral, pais, estudantes, professores e administradores, bem como ao ministro da educação. A agência também desenvolveu uma ampla variedade de materiais opcionais de ensino e aprendizado especialmente desenhados levando em consideração o desempenho dos estudantes e os pontos fortes e fracos dos professores. Assim como diversas autoridades dos Estados Unidos, a DEP procurou classificar o desempenho das escolas pelos padrões internacionais.

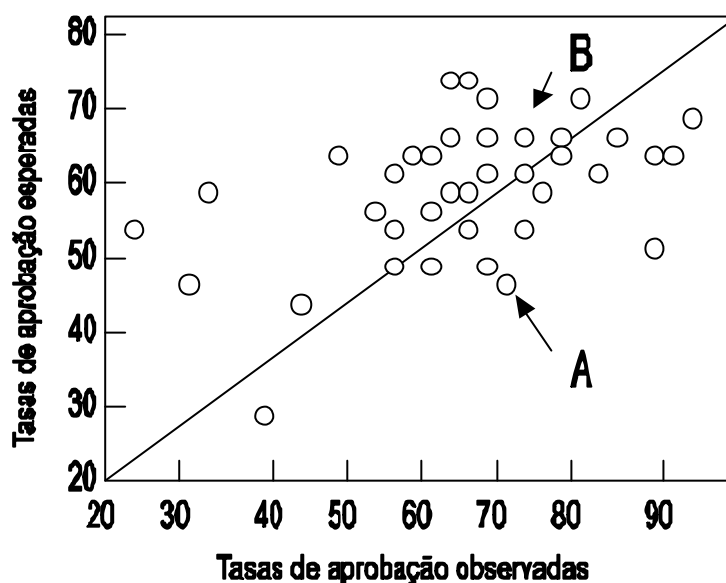
Na França, como no mundo inteiro, as escolas das comunidades mais pobres alcançam pontos mais baixos que as mais ricas. A DEP procurou identificar escolas em áreas atingidas pela pobreza que estejam tendo bom desempenho para melhor definir o “valor agregado” da escolarização. Uma abordagem estatística à classificação do desempenho das escolas em relação ao status socioeconômico de seus estudantes é mostrada no Gráfico 1. As escolas com desempenho melhor

do que se poderia esperar de sua população estudantil são denominadas escolas “eficazes”. Estudando tais escolas, as autoridades deverão ser capazes de determinar por que elas são eficazes, replicando então essas condições nas

escolas de desempenho mais fraco. Este tipo de análise já vem sendo feito em algumas partes dos Estados Unidos (ver, por exemplo, a descrição do sistema de “valor agregado” do sistema escolar de Dallas relatada *in* Alvarez e Ruiz-Casares, 1997).

Gráfico 1

Aferindo o valor agregado em educação: um exemplo da França Escolas “eficazes” produzem resultados melhores que os esperados de uma dada população estudantil



Este gráfico mostra um método utilizado pelo governo francês para classificar o desempenho das escolas em relação ao status socioeconômico de seus estudantes. Cada ponto representa uma escola, plotado de modo a representar as taxas reais (verificadas) e esperadas de aprovação no exame de conclusão (*baccalauréat*). Os alunos das escolas localizadas abaixo da linha têm maior chance de passar no exame, comparado com o que seria de se esperar em função de sua idade e origem social.

A escola A teve valor agregado em termos de sucesso no exame porque sua taxa verificada é maior que a esperada (está localizada abaixo da linha). Este não é o caso da escola B, que não alcançou o nível de sucesso esperado, apesar de sua taxa de aprovação verificada ser, na verdade, maior que a de A.

Fonte: adaptado de “*L'état de l'école*”, no. 5, outubro de 1995. *Direction de L'Évaluation et de la Prospective (DEP)*, *Ministère de L'Éducation Nationale, de L'Enseignement Supérieur, de la Recherche, et de l'Insertion Professionnelle (MEN)*, p. 51. Os dados usados na preparação desta ilustração não correspondem necessariamente a resultados reais de testes.

Países em Desenvolvimento

Nos países em desenvolvimento, há também um movimento em prol do estabelecimento de sistemas nacionais de avaliação da educação. Fora da América Latina, foram descritos e analisados sistemas de avaliação na Tailândia, no Egito, na Coreia do Sul e na Jordânia, entre outros. Embora seja pequena e pobre, a Jordânia desenvolveu um programa de avaliação sistemático e tecnicamente competente. Seu sucesso neste aspecto pode ser atribuído: a) ao envolvimento direto das mais altas autoridades (o príncipe herdeiro, por exemplo) e seu compromisso de longo prazo com a melhoria da qualidade da educação e a comunicação aberta dos resultados; b) ao estabelecimento de uma agência responsável pela avaliação independente, bem dotada de recursos financeiros e fora do Ministério da Educação; c) à forte liderança, aliada à assessoria do exterior. Em 1991 a Jordânia realizou diversas rodadas de testes baseados nas avaliações internacionais desenvolvidas pelo teste IAEP do Educational Testing Service e hoje presta assessoria técnica nesta área a

países vizinhos. Uma rodada de avaliações recentemente concluída verificou melhoramentos significativos no aprendizado dos alunos, resultado dos esforços jordanianos, que duram mais de uma década, para reformar os currículos escolares e o treinamento dos professores.

III A EXPERIÊNCIA LATINO-AMERICANA

Até 1991 o Chile e o México eram os únicos países da região com relevante experiência em avaliações. O programa chileno de avaliação está em operação desde 1980 e o do México, apesar de implantado há vários anos, teve seus resultados pouco difundidos porque as autoridades pouco fizeram para divulgá-los. A Costa Rica realizou avaliações nacionais entre 1986 e 1990. A Colômbia possuía há muito um sistema nacional de testes, utilizado principalmente para seleção na admissão à educação superior.

Nos anos imediatamente seguintes, outros países latino-americanos começaram também a mover-se nesta direção. Por volta de 1996 a experiência havia sido consideravelmente ampliada e quase todos os países da América Latina

havam iniciado algum tipo de programa de avaliação. Este relatório tira suas conclusões do que ocorreu no Chile, na Costa Rica, no México, na Argentina, no Brasil e na Colômbia. Descreve ainda o programa regional da UNESCO/Orealc.

Chile

Como já foi mencionado, há muito que o Chile desenvolve uma avaliação educacional. Seu programa foi concebido em 1978, quando o Ministério da Educação solicitou à Pontifícia Universidad Católica que desenhasse e implementasse um sistema de informações para a educação. Em 1988, com a transferência das escolas públicas para os municípios, o nome do programa foi alterado para Programa para Aferição da Qualidade da Educação Básica no Chile (SIMCE).

A função do SIMCE é auxiliar o Ministério da Educação e as autoridades provinciais na supervisão do sistema educacional, na avaliação das escolas individuais e na assistência ao treinamento dos professores no trabalho. O programa testa crianças nas quarta e oitava séries em espanhol e aritmética, e 10 % delas em ciências naturais, história

e geografia. Avalia, ainda, desenvolvimento pessoal e atitudes, as atitudes e o *background* de professores e pais, e a eficiência da escola. As avaliações nessas duas séries ocorrem em anos alternados. A partir de 1991 o Ministério da Educação assumiu integralmente a administração do programa.

A partir de 1988 o programa tornou-se mais eficiente e eficaz, após melhoria da capacidade técnica, da informatização e da administração. As notas são agora entregues às escolas com maior rapidez e os relatórios foram simplificados para assegurar que os resultados sejam facilmente compreendidos. Em muitas escolas, os esforços parecem estar influenciando no planejamento pedagógico. Os encarregados do desenho dos currículos e dos materiais de instrução, por exemplo, estão dando ênfase ao domínio das matérias toda vez que os alunos apresentam problemas. Alguns pais estão usando os resultados do SIMCE para escolher as escolas de melhor desempenho para seus filhos. O custo de programa é cerca de US\$5,00 por estudante, comparável aos padrões internacionais.

As notas do SIMCE também formaram a base de um programa pedagógico inicialmente dirigido para as 900 – e posteriormente 1.200 – escolas chilenas de desempenho mais fraco. Sob este programa, as escolas receberam materiais educacionais, bibliotecas, livros, infra-estrutura e treinamento no serviço. Dependendo de suas condições iniciais, certas escolas têm preferência na alocação de subsídios para programas de melhoria local. Além disso, as escolas cujas notas aumentam de um ano para outro recebem recompensas financeiras.

Os testes do SIMCE revelaram o seguinte: as escolas com alunos oriundos de famílias pobres e não educadas, ou das áreas rurais, têm as piores notas nos testes do SIMCE; as escolas públicas municipais e as escolas rurais têm pior pontuação que as escolas privadas, piores especialmente que as de instituições estabelecidas de longa data; as escolas privadas têm desempenho melhor que as escolas públicas, mesmo depois de feito controle para compensar o status socioeconômico dos pais; e, finalmente, os pontos das 900 escolas parecem ter melhorado significativamente nos últimos anos.

Os problemas atuais incluem o fato de que, conquanto muitas escolas estejam utilizando os resultados do SIMCE para aprimorar as condições locais, esperam ainda que as ações corretivas sejam iniciadas pelas autoridades centrais. Algumas das escolas informaram um aumento no número de alunos carentes, em um aparente esforço para mostrar que seus ganhos relativos (valor agregado) melhoraram. As aferições do campo afetivo não foram bem sucedidas e devem talvez ser abandonadas. Pode ser apropriado, ainda, reduzir o tamanho do universo que está sendo testado, utilizando-se mais a amostragem. Finalmente, devido a problemas técnicos, a comparação dos resultados de um ano para outro é inadequada.

De um modo geral, o Chile tem agora o sistema de avaliação mais compreensivo e melhor administrado da América Latina, e o SIMCE serviu como poderoso instrumento para implantação de um programa de reforma, incentivando a descentralização, a prestação de contas (*accountability*), e maior aprendizado. As autoridades governamentais estão procurando melhorar ainda mais o sistema. Planejam agora, em particular,

adicionar às avaliações um teste de desempenho (atualmente todos os testes são do tipo múltipla escolha). Uma vez que as causas do desempenho fraco ainda não estão completamente entendidas, um esforço mais sofisticado de pesquisa está em andamento. Finalmente, as autoridades chilenas decidiram participar do *follow-up* do estudo do TIMSS previsto para 1999, como forma de estabelecer os padrões de comparação de seus estudantes com os do resto do mundo.

Costa Rica

A Costa Rica realiza avaliações educacionais desde 1986. Sua atividade nesta área avançou em três estágios. Durante o primeiro deles – de 1986 a 1990 – os objetivos estabelecidos foram: aferir até que ponto as crianças e os jovens estavam aprendendo conceitos básicos; encorajar pais e professores a utilizar seu tempo de ensino com maior eficácia; estimular uma discussão nacional sobre a qualidade da educação; sinalizar que todos os costarriquenses eram responsáveis pela melhoria da qualidade da educação; e demonstrar a importância do restabelecimento de exames

nacionais de certificação na conclusão do secundário.

O programa foi, em grande parte, formulado pelo IIMEC (Instituto de Pesquisa para Melhoria da Educação na Costa Rica), uma instituição autônoma da Universidade da Costa Rica. Inicialmente, o programa introduziu testes de espanhol e matemática para o universo dos alunos das 3^a, 6^a, 9^a, 11^a e 12^a séries. Em 1987 e 1988 os testes foram ampliados, passando a incluir ciências físicas, ciências sociais, inglês e francês. Em 1988 foi instituído um exame de conclusão do secundário. Esses testes levaram a uma controvérsia pública, devido ao baixo nível de desempenho dos estudantes, e os testes objetivos abrangentes foram reduzidos.

De 1988 a 1993 o foco principal foi para o exame de conclusão do secundário. Os testes, preparados pelo IIMEC e implementados pelo Ministério da Educação, valiam 60% dos pontos finais de cada estudante e as correções eram feitas no nível local. Quando um estudo de 1991 revelou que as notas de 30% dos testes estavam incorretas, a favor do estudante em quase todos os casos,

as autoridades decidiram eliminar a maioria das questões de desempenho e centralizar a correção de praticamente todos os testes, utilizando leitura ótica. Além disso, os exames da 6ª série, preparados pelas autoridades regionais, valiam 50% das notas finais dos estudantes, mas pouco esforço foi feito para garantir que os resultados informados pelas autoridades regionais fossem confiáveis, ou válidos.

A partir de 1993, sob um novo governo e com novo ministro da educação, a Costa Rica comprometeu-se firmemente com a avaliação, recebendo apoio através de um projeto financiado pelo Banco Mundial. Testes de diagnóstico em quatro matérias foram restabelecidos para as 3ª, 6ª e 9ª séries; outros tipos de avaliação foram implantados; e o IIMEC recebeu reforços em equipamentos e pessoal. O novo sistema administra uma ampla variedade de testes e avaliações, a maioria preparada pelo IIMEC.

Entre as avaliações formativas estão:

1.) *Testes de desempenho para diagnóstico* nas três matérias básicas da 3ª série. Foi feito um

ensaio prévio deste teste em 1995 e em 1996 foi aplicado a uma grande amostra.

2.) *Avaliações de diagnóstico inicial* aplicadas a 10% das crianças que entram na 1ª série, para obter informações sobre seus status físico, cognitivo e socioemocional. Os resultados serão usados para estabelecer linhas mestras para introdução de experiências apropriadas de aprendizado no pré-escolar e na 1ª série.

3.) *Testes de habilidade na solução de problemas*, aplicados em 1996 a uma amostra de alunos da 9ª série. Foram desenhados para aferir as capacidades cognitivas nesta área, bem como as variáveis socioeconômicas e acadêmicas que influenciam o desempenho.

4.) *Avaliação da capacidade física*, realizada em 1996 para aferir a aptidão física de uma amostra nacional de crianças das 3ª, 6ª e 9ª séries.

Diversas avaliações somativas foram também realizadas. Um exame de conclusão preparado no nível nacional é aplicado a todos os alunos da 9ª série, em todas as matérias

básicas. As questões de múltipla escolha são corrigidas por leitura ótica, enquanto as notas das questões de desempenho são avaliadas por professores especialmente treinados. Os resultados valem 25% dos pontos do segundo semestre. Desde 1988 a Costa Rica também aplica um *exame de conclusão da escola secundária*, que é elaborado e aplicado regionalmente. Este exame vale 60% da qualificação final e compreende tanto questões objetivas quanto de desempenho. Além deste, exames de conclusão da 6ª série são preparados e aplicados pelos escritórios regionais do Ministério da Educação, mas variam em qualidade, validade e confiabilidade.

Nos testes iniciais (1988), as crianças tiveram desempenho muito abaixo das expectativas do currículo nacional. As escolas urbanas privadas tiveram melhor desempenho nesses exames. Apesar do IIMEC distribuir informações resumidas a todas as escolas, as instituições não fizeram nenhuma tentativa explícita de utilizar os resultados. Contudo, uma pesquisa entre os professores, de 1989, indicou que 70% deles tinha conhecimento do material e 35% o havia, de algum modo, usado nas salas de aulas.

No segundo estágio, as informações dos exames de conclusão da escola secundária foram postas à disposição dos estudantes e escolas secundárias individuais. Como utilizar tais informações, foi deixado inteiramente a critério da escola.

Espera-se que o terceiro estágio demonstre uma abordagem mais sistemática à divulgação, utilização e *feedback*. Serão preparados relatórios para diferentes níveis de séries, e pedir-se-á às unidades regionais de educação que incorporem os resultados da avaliação em seus planos educacionais. Será feita, ainda, tentativa de aferir a extensão da melhoria da educação. Apesar desses progressos: os pais não recebem qualquer informação além das notas do estudante; não se distribuem à imprensa notas sobre educação; os resultados não são utilizados para preparar materiais específicos que corrijam as deficiências no aprendizado; e as instituições de treinamento não recebem qualquer orientação ou informação específicas sobre os testes. São, contudo, fornecidos relatórios às escolas individuais, especialmente os que se referem às atividades do pré-escolar e da 1ª série,

acompanhados de sugestões para enriquecimento do currículo.

De um modo geral, os esforços de avaliação na Costa Rica foram prejudicados por uma abordagem cheia de interrupções e reinícios. Embora o ministro da educação anterior tenha dado amplo apoio ao programa, alguns educadores e líderes políticos não o fizeram, e não está claro se esta situação irá continuar sob a nova administração empossada em maio de 1998. Dado seu pequeno tamanho, a Costa Rica pode bem ter embarcado em um programa demasiadamente ambicioso. Um curso de ação mais produtivo pode ser reduzir o número de testes e concentrar-se na divulgação e utilização dos resultados. O exame costarricense de “grandes ganhos” (*high stakes*) da escola secundária, que vale 25% da nota final e é utilizado como meio para elevar o desempenho neste nível, pode ser de interesse para outros países latino-americanos.

Colômbia

Em 1986 o governo da Colômbia criou um departamento no Ministério da Educação que ficou responsável pela avaliação e aferição

das instituições e programas educacionais. Em 1990, iniciou um sistema de avaliação nacional com vistas à construção de um modelo operacional de avaliação que levasse a decisões que melhorassem a qualidade. Mais ainda, esperava-se que o sistema avaliasse o conhecimento dos estudantes e determinasse como professores, escolas e materiais educacionais promoveriam este conhecimento; que gerasse um maior senso da importância da escolarização; que comunicasse este senso a toda a sociedade; que fornecesse base teórica para conceitos como qualidade e avaliação de desempenho educacional; e que desse suporte à pesquisa que iria ajudar o país a alcançar esses objetivos. O programa foi realizado por uma equipe interinstitucional composta pelo pessoal do Ministério da Educação, da Universidade Nacional de Professores, do Centro de Estudos Sociais e do Instituto SER.

Desde 1968, o Instituto Colombiano de Aprimoramento da Educação Superior (ICFES) vem sendo responsável pela formulação de um exame nacional para admissão às instituições de ensino superior do país. Em 1980 tais exames foram tornados

obrigatórios e passaram a ser aplicados pelo Serviço Nacional de Testes do ICFES. Conquanto seu enfoque seja a avaliação do estudante e sua admissão às instituições de ensino superior, em vez de visar a avaliação do sistema educacional como um todo, ou de seus subsistemas, o programa ajudou a criar competência técnica na elaboração e aplicação de testes.

Ao contrário do Chile e da Costa Rica, a Colômbia enfatiza a amostragem e a pesquisa como meios de identificar as causas do baixo desempenho, o que leva a que as informações sobre desempenho não sejam fornecidas a todos os pais, alunos ou escolas. No momento, a expansão da amostra está sendo contemplada para assegurar que um maior número de municípios e escolas receba *feedback*.

O sistema colombiano iniciou-se com um teste de matemática e espanhol para alunos da 3^a e 5^a séries. Foram subseqüentemente desenvolvidos testes em ciências naturais e sociais para as 7^a e 9^a séries. Esses testes mediam explicitamente a capacidade de raciocínio de ordem mais elevada, tal como o uso de algoritmos e habilidade

na solução de problemas em matemática, e a interpretação de textos em espanhol.

Quando os desempenhos cognitivos dos alunos das 3^a e 4^a séries são medidos em confronto com o currículo nacional, os resultados são desapontadores. Poucos estudantes são capazes de efetuar as operações básicas exigidas na solução de problemas concretos e poucos compreendem inteiramente o que lêem, seja em sentido crítico, seja reflexivamente. As notas mais altas são obtidas nas escolas urbanas e privadas, e em determinadas regiões. Mesmo em um estudo em que se estabeleceu controle do status socioeconômico, verificou-se que freqüentar escolas privadas parece assegurar vantagens em termos de desempenho. O desempenho foi melhor entre os que haviam freqüentado o pré-escolar, os que não haviam repetido ano, os que raramente faltavam, os que possuíam livros em casa e tinham pais com níveis mais elevados de educação. Se havia livros textos disponíveis na escola, e se eram usados, o desempenho era ainda melhor. Os alunos com professores mais bem treinados, com mais livros textos, com professoras (mulheres) e que

freqüentavam escolas completas também tiveram bom desempenho. Os testes confirmaram ainda que os estudantes da *Escuela Nueva*, um programa inovador para pequenas escolas rurais, apresentavam progresso significativamente maior que os alunos das escolas rurais que não estavam no programa.

As conclusões dessas avaliações influenciaram os objetivos e o conteúdo da nova lei de educação aprovada em 1994. Contudo, as informações reunidas através das atuais avaliações baseadas em amostragem são de pouca utilidade para escolas e municípios específicos, que se tornaram progressivamente mais responsáveis por suas escolas sob o novo sistema educacional descentralizado. Embora as autoridades atuais sejam de opinião que uma expansão da avaliação a todos os estudantes seria demasiado onerosa, a Colômbia pretende trabalhar com amostras maiores, e espera estabelecer um banco de dados e questões que possa ser usado diretamente pelas escolas e autoridades educacionais.

Como já foi mencionado, a Colômbia é o único país da América Latina que participou integralmente do

programa do TIMSS. Ainda que os resultados tenham sido desapontadores, merece ser aplaudida por sua decisão de aceitar o escrutínio externo. Ela busca agora utilizar os resultados do TIMSS para reforçar o sistema nacional de avaliações e reformar o currículo.

O programa colombiano de avaliação tem um forte elemento de pesquisa que afetou a política global, notadamente na redação da lei de educação de 1994, e confirmou o sucesso do *Escuela Nueva*. Os resultados, porém, ainda não foram utilizados sistematicamente para aprimorar o desempenho de escolas individuais, ou para reformar os currículos. O currículo de matemática e ciências, em particular, precisa ser revisto. Paralelamente, a parceria entre instituições públicas e privadas vem funcionando bem.

Brasil

O Brasil está entre os países que iniciaram tardiamente seus instrumentos de avaliação em nível nacional. Ele estabeleceu o Sistema Nacional de Avaliação da Educação Básica (SAEB) em 1990 e só começou a fazer amostragem em 1993 e 1995. Um dos objetivos do SAEB tem sido

encorajar os estados e municípios a instituírem suas próprias avaliações. Os estados do Paraná, Minas Gerais e o Distrito Federal de Brasília particularmente, recentemente iniciaram programas de avaliação.

Em 1995 o SAEB introduziu várias inovações: sua amostragem passou a incluir o ensino secundário e escolas privadas; adotou métodos mais sofisticados de aferição e introduziu instrumentos que fornecerão informações sobre o *background* dos estudantes e reduzirão o tempo para publicação dos resultados. A pesquisa de 1995 enfocou as 4^a e 8^a séries da educação básica e a 3^a série do nível secundário.

Os resultados de pesquisa publicados em 1995 indicam que 90.949 alunos foram submetidos a exames nas 4^a e 8^a séries do primário e nas 1^a e 2^a do secundário. Esta amostra de estudantes foi tirada de 2.289 escolas públicas e 511 privadas. Os itens da pesquisa foram baseados no nível de aprendizado esperado pelos professores e especialistas em educação. Em leitura, os estudantes foram testados em compreensão, extensão e exame crítico de conteúdo. Em matemática, a pesquisa

concentrou-se em três categorias: compreensão de conceitos, entendimento e aplicação de procedimentos e solução de problemas. Na análise estatística, a principal tarefa foi a determinação dos níveis esperados e reais de desempenho das crianças nas várias séries.

Nos testes de 1995, os estudantes em todo o país tiveram resultados significativamente inferiores ao níveis esperados pelos professores e especialistas. Em matemática, apenas 21 por cento dos alunos da 4^a série pontuaram acima do nível esperado; na 8^a série, foram apenas 15 %, e 4% no secundário. As notas em língua foram ainda menores: apenas 22 % dos alunos da 4^a série, 14% dos da 8^a, e 1% dos secundaristas pontuaram acima do nível esperado. As notas mais altas vieram das regiões Sul, Sudeste e Centro-Oeste do país. As menores notas foram encontradas nas regiões Norte e Nordeste. Os estudantes das principais cidades tiveram melhores resultados que os do interior; as crianças com pais de melhor nível de educação, e as que freqüentavam escolas privadas também tiveram pontuação mais elevada; os estudantes mais velhos tiveram

melhor desempenho que os mais jovens; os estudantes brancos e asiáticos saíram-se melhor que os mestiços e pretos; e os estudantes com professores mais bem treinados tiveram, em geral, notas mais altas que aqueles com professores menos bem treinados. Estes resultados contêm implicações importantes para a política educacional, mas as autoridades mal começaram a discutí-los. Mais, precisam ainda incorporar técnicas mais sofisticadas em suas análises, tais como a introdução de controle para o status socioeconômico dos estudantes que ingressam.

Recentemente, o Estado de São Paulo deu início a um amplo programa de avaliação. Todos os alunos em séries selecionadas do primário e do secundário são submetidos a exames de matemática e língua, e os resultados são enviados aos pais, aos professores e às escolas. É fornecida uma análise dos itens onde os estudantes tiveram pior desempenho. “Itens âncora” são incluídos, para assegurar a compatibilidade de resultado de ano para ano. O programa de São Paulo é executado principalmente através da contratação de agências privadas, sem fins lucrativos, de elaboração e aplicação de testes. O programa de

avaliação de Minas Gerais também parece estar apresentando progressos na utilização de seus resultados desde que os pontos individuais passaram a ser informados às escolas.

De modo bem independente do que foi relatado acima, mediante uma lei aprovada em 1995, o Ministério da Educação desenvolveu um sistema para avaliação das instituições de ensino superior. As escolas com baixos resultados terão de elaborar programas para reforço de qualidade, que terão o apoio do governo federal.

Em resumo, o programa de avaliação brasileiro apenas recentemente foi iniciado. Só agora as autoridades brasileiras começam a considerar como as avaliações podem ser utilizadas para aprimorar a política educacional e o currículo. Há necessidade de realizar análises mais rigorosas para identificar os fatores que afetam o desempenho. Uma vez que a educação no Brasil é bastante descentralizada e não segue um currículo nacional, o governo está contemplando a implantação de algum tipo de padrões nacionais voluntários e um exame, também nacional, para verificar se tais padrões teriam sido atingidos. O programa de avaliação

do ensino superior, o primeiro deste tipo na região, é uma inovação importante.

México

Em 1970, o México estabeleceu um departamento na Unidade de Planejamento da Educação, da Secretaria de Educação Pública, e que acabou transformando-se na Subdiretoria de Avaliação e Diplomação, para examinar as características e a qualidade do sistema educacional do país. Subseqüentemente, a equipe aplicou testes de aptidão às crianças da sexta série do primário, e estabeleceu um exame de admissão às escolas secundárias.

De 1976 a 1982, a subdiretoria investigou o nível de aprendizado de uma amostra representativa de alunos da 4ª e 5ª séries. Os resultados desta avaliação apareceram em publicações científicas e acadêmicas, mas não tiveram qualquer outra divulgação, e as autoridades deram-lhes pouca atenção. Na verdade, as informações da avaliação tornaram-se um "segredo de Estado", conhecidas apenas por um pequeno número de funcionários da secretaria. Este comportamento

prejudicou o desenvolvimento técnico e a utilização nas políticas do setor.

Durante o período de 1983 a 1988, o México desenvolveu um exame para os formados nas escolas de treinamento de professores. Em 1989 decidiu aplicar mais amplamente o conceito de avaliação, visando melhorar o ensino e o aprendizado, e divulgar os resultados. Em 1992, o governo federal e os Sindicatos Nacionais de Professores chegaram a um acordo sobre um programa para modernização da educação básica através da descentralização para o nível estadual, mas mantendo no governo federal a função de aferir e avaliar o aprendizado e de garantir a qualidade da educação básica e do treinamento dos professores. Para tanto, a Secretaria de Educação Pública comprometeu-se a apoiar avaliações nacionais, de professores e de salas de aulas.

Em 1994, após cinco anos de avaliação da qualidade da educação no México, a secretaria divulgou um relatório sobre a competência e habilidade de 480.000 professores e sobre o desempenho de 2,8 milhões de alunos do primário e do secundário. Suas principais conclusões foram que os estudantes que haviam freqüentado

o pré-escolar tiraram notas mais elevadas que os que não o haviam feito; que as crianças que repetiam a sexta série, ou que trabalhavam, tiveram pior desempenho que seus colegas; que as crianças que freqüentavam escolas urbanas, ou privadas, tiveram desempenho muito superior ao daquelas das escolas rurais, ou públicas; as que tiveram o pior desempenho foram as de escolas indígenas ou comunitárias, com instalações pobres e professores menos bem treinados; e as que tiveram o melhor desempenho foram as que estudavam em escolas urbanas e tinham pais com melhor educação. Embora as crianças das primeira e segunda séries apresentassem resultados mais próximos do que esperavam os pesquisadores e os elaboradores de currículos, suas notas, como percentual de respostas corretas, caíram nos anos subseqüentes. As autoridades mexicanas informaram também que era impossível aferir sistematicamente os desempenhos dos professores nas salas de aulas, porque as populações de alunos ainda são extremamente distintas, e as dificuldades técnicas ainda abundam.

Como mencionado anteriormente, o México participou

integralmente no TIMSS, mas no último momento decidiu não divulgar os resultados. Esta decisão não deveria causar tanta surpresa, em vista da relutância do México em revelar os resultados de exames no passado, resultados estes que às vezes foram tratados como segredos de Estado. Apesar de haver mudado a atitude desde então, e de que os resultados das avaliações do primário e do secundário são agora divulgados publicamente, a informação ainda é dada com certa ambivalência. A característica mais distintiva do sistema de avaliação mexicano é que ele vem testando sistematicamente o conhecimento e a capacidade dos professores.

Argentina

Assim como o Brasil, a Argentina começou tarde seu programa de avaliações, mas agiu com maior rapidez no estabelecimento e utilização de padrões para melhoria da qualidade da educação. Sob uma nova lei federal que descentraliza o ensino, o Ministério da Educação estabeleceu, em 1993, um sistema de avaliação nacional, dentro da Secretaria de Programação e Avaliação Educacional. Os objetivos do sistema de avaliação são:

promover a descentralização; fornecer informações fundamentais sobre o estado da educação; acompanhar o progresso em direção aos objetivos da reforma; identificar desigualdades e inadequações e áreas em que sejam necessários programas compensatórios para as populações em desvantagem; e encorajar setores mais amplos a participarem no processo decisório da educação. Desde 1993, a Argentina aplicou testes a todas as crianças no último ano do primário e do secundário, em matemática e língua. Em 1995 as avaliações foram estendidas à terceira série do primário e à segunda do secundário. Foram iniciados, também, testes em ciências naturais e sociais. Questionários complementares foram dados a professores, diretores, estudantes e pais de alunos. Os testes basearam-se em uma análise detalhada do aproveitamento curricular esperado e cuidadosos testes-piloto. A confiabilidade e a validade desses testes foram confirmadas através de análise estatística e de revisão por especialistas da correspondência entre os itens dos testes e os objetivos do currículo. O programa é integralmente administrado pelo Ministério da Educação. Os

resultados das avaliações são incorporados em um sistema mais abrangente de informações sobre a educação que é usado para monitorar e supervisionar a educação em todos os níveis.

As informações das avaliações incluem detalhes sobre desempenho das escolas e perfil dos estudantes e professores, além de modelos de gerência associados a desempenho. Estas informações são divulgadas às autoridades nacionais, estaduais, locais e escolares. A pontuação das escolas individuais, contudo, não é informada às escolas. Diversos relatórios tratam, ainda, do conteúdo de matéria e incluem recomendações para aprimoramento da pedagogia, tanto em nível central, quanto no das escolas. Os resultados dos testes foram também usados para desenvolver manuais sobre como melhorar o aprendizado, bem como para treinamento no serviço, e tais guias, junto com assistência técnica, foram fornecidos a escolas em todas as 24 províncias do país.

Embora as perguntas tivessem sido elaboradas para um nível mínimo esperado de respostas, na média, os estudantes responderam apenas 50 por cento das perguntas corretamente.

Fatores ligados às escolas justificam 40 por cento das variações associadas com as notas de matemática e 28 por cento das referente à língua. As crianças oriundas das famílias no cume da escada socioeconômica, tiraram as notas mais elevadas. Contudo, as diferenças entre escolas foram maiores que as diferenças dentro da escola, o que sugere que o sistema argentino de ensino é altamente segmentado. As crianças pobres em turmas heterogêneas tiraram notas melhores que as crianças pobres em turmas homogêneas. De um modo geral, os resultados indicam que há espaço para uma significativa melhora dentro das escolas.

Apesar de ter sido iniciado apenas em 1993, o programa argentino de avaliação é bem planejado e estreitamente vinculado à sua política de descentralização e melhoria da qualidade da educação em geral. Como o Chile, a Argentina está determinada a utilizar os resultados de suas avaliações. Já foi feita ampla divulgação dos relatórios nacionais e regionais, muitos dos quais abordam problemas de grande importância, além de fornecerem materiais desenhados para reforçar o ensino.

Programa de Avaliação Regional da UNESCO/Orealc

Através de uma dotação do BID, o escritório regional da UNESCO (*Oficina Regional para América Latina y el Caribe, Orealc*) está implementando um programa de avaliação chamado “laboratório latino-americano de aferição da qualidade da educação”. O programa, que foi iniciado em 1995, busca encorajar a coordenação entre agências nacionais de avaliação da região, e reforçar sua capacidade. Em 1997, testes de matemática e leitura, acompanhados por questionários de *background* detalhados, foram aplicados em quinze países latino-americanos a uma amostra de alunos da terceira e quarta séries. Os testes foram elaborados por um comitê composto por todos os países participantes e basearam-se em uma revisão informal dos objetivos dos currículos de matemática e leitura da região. Os resultados serão informados por volta de meados de 1998.

Em 1993, a UNESCO implantou um teste similar, em base piloto, aplicado aos alunos de quarta série de sete países latino-americanos. A maioria dos estudantes

teve pontuação muito abaixo dos níveis mínimos esperados pelos especialistas regionais em currículo e elaboração e aplicação de testes. Na média, os estudantes foram capazes de responder cerca de metade das

questões corretamente. A Venezuela e a Costa Rica tiveram as pontuações mais elevadas, enquanto a República Dominicana, a Bolívia e o Equador ficaram com as mais baixas. A Tabela 5 resume os resultados dos testes.

Tabela 5

Pontuação nos Exames Regionais de Matemática e Língua da Quarta Série, 1994		
Percentagem de Alunos que Respondeu Corretamente Três Quartos ou Mais das Perguntas		
<u>País</u>	<u>Matemática</u>	<u>Língua</u>
Venezuela	15,5	25,2
Costa Rica	10,4	23,8
Chile	5,0	23,2
Argentina	1,3	18,5
República Dominicana	1,4	5,0
Bolívia	2,1	4,9
Equador	0,4	7,7
Região	5,5	15,8

Fonte: UNESCO/Orealc, "Medición de la Calidad de la Educación: Resultados", (Santiago do Chile, 1994)

A UNESCO está planejando encomendar um estudo comparando o grau de dificuldade e o conteúdo curricular de seus testes de matemática com os do TIMSS. Isto dará uma evidência direta da qualidade do desempenho nos países da América Latina em comparação com outros países fora da região. Os especialistas do TIMSS poderiam também prestar assistência a países latino-americanos selecionados na implantação da metodologia do TIMSS

para aferição da "oportunidade de aprendizado" – revisando e comparando, por exemplo, currículos oficiais, currículos implementados, livros-textos, e pedagogia de sala de aulas.

O programa da UNESCO é um passo importante em direção ao estabelecimento de uma coordenação regional. Para assegurar o fortalecimento das capacidades regionais no longo prazo, os esforços

futuros deverão apoiar, mais que no passado, os centros regionais de excelência, vinculando instituições públicas e privadas, e mantendo laços mais fortes com a IEA e outros centros mundiais avançados de pesquisa e desenvolvimento.

IV – LIÇÕES APRENDIDAS E DESAFIOS FUTUROS

Resumo da Experiência em Seis Países

O Chile e a Argentina têm os sistemas de avaliação mais abrangentes e mais bem administrados da América Latina. O Chile demonstrou, também, o mais forte compromisso de longo prazo com as avaliações. Embora Chile e Argentina tenham dado os passos mais largos na utilização das avaliações para fins de política, reforma curricular e melhoria em escolas individuais, mesmo nestes países há ainda muito campo para aprimoramento. De modo particular, ainda não foram estabelecidos objetivos claros de aprendizado, e esforços para “alinhamento” de currículos, livros-textos e pedagogia de sala de aulas apenas começaram. O Chile e a Argentina fizeram algum progresso também na integração dos

sistemas de avaliação a grandes sistemas de monitoração e avaliação.

Entre os problemas-chave que exigem maior atenção estão: a relutância do México em divulgar os resultados das avaliações; a abordagem de arrancos e paradas da Costa Rica; e o lento progresso do Brasil na utilização das avaliações para reforma de currículo e políticas. Com a exceção da Costa Rica, os seis países discutidos aqui preferiram basear-se em testes de múltipla escolha, em vez de em testes baseados em desempenho ou em perguntas abertas.

Outro ponto a observar é que todos os seis países informaram resultados similares de avaliação. De modo particular, os estudantes das últimas séries do primário e do secundário tiveram pontuações muito abaixo das expectativas dos educadores e pesquisadores profissionais. Os estudantes das escolas urbanas e privadas, e com pais mais bem educados, tiveram as melhores pontuações. Poucos países da região tentaram até o momento realizar análises multivariadas detalhadas desses resultados, para determinar suas causas. Embora diversos estudos tenham revelado

uma correlação entre professores mais bem treinados e melhor desempenho dos alunos, a maioria não fez qualquer controle adequado para status socioeconômico dos estudantes, ou localização da escola, para confirmar a importância da educação do professor. Além disso, pouco se sabe ainda sobre a mudança do desempenho dos estudantes ao longo do tempo.

Entre as inovações dignas de nota na região, estão: a tentativa brasileira de tentar avaliar as instituições de ensino superior; as avaliações, na Costa Rica, do preparo para o aprendizado das crianças que estão sendo admitidas às escolas primárias e seu exame dos estudantes que estão completando o secundário; os testes mexicanos de conhecimento e competência dos professores; a ênfase em pesquisa da Colômbia; a utilização das avaliações para orientar a alocação de recursos, no Chile; e os esforços argentinos para utilizar as avaliações na reforma do currículo. O programa da UNESCO/Orealc também é uma inovação importante, que terá maior impacto à medida que crescer a coordenação com instituições como a IEA.

Construção de Consenso e Compromisso

Construção de Consenso Nacional

A lição mais importante até o momento talvez seja que os países da América Latina já não precisam discutir se irão ou não realizar avaliações. Em vez disso, precisam estabelecer metas educacionais, verificar se as crianças, as instituições e os sistemas escolares estão atingindo essas metas, e iniciar programas que garantam que tais metas sejam eventualmente alcançadas. As avaliações não ajudarão a melhorar a qualidade da educação a menos que todos concordem a respeito da importância de se melhorar esta qualidade, e que haja divulgação integral dos resultados a todos os interessados, de forma tempestiva e que permita uma fácil leitura (isto é, que haja “transparência”). Este consenso precisa representar uma coalizão de professores, pais, administradores e líderes empresariais e políticos. Precisa ter continuidade no longo prazo e contar com o apoio das mais altas autoridades governamentais, mas suas decisões não podem ser

afetadas por partidarismo político. Os programas de avaliação precisam avançar sem paradas e reinícios intermitentes (como na Costa Rica) ou sem que seus resultados sejam ocultados do público (como no México).

Integrando os Professores no Processo de Avaliação

Para que as avaliações tenham impacto na educação, é preciso que os professores sejam incluídos no processo, desde o início. Os sindicatos de professores em toda a América Latina tendem a ser observadores passivos, ou mesmo opositores da avaliação. É preciso persuadi-los de que é melhor para eles próprios assumir uma perspectiva mais profissional em relação a seu trabalho, aprimorar os procedimentos de sala de aula e participar integralmente em todos os programas de avaliação. Os sindicatos não devem recear que os resultados possam ser usados para premiar ou punir professores individuais, já que o corpo docente varia tanto de ano para ano que será sempre difícil aferir com precisão a qualidade do ensino, como o indicam as experiências da França, dos Estados Unidos e do México. Se as avaliações forem vistas como um

meio de avaliar as escolas e os sistemas escolares locais, em vez de professores, então pais, professores e administradores terão maior possibilidade de trabalhar em conjunto para melhorar os índices de desempenho das escolas.

Uma Abordagem Sistêmica

Um compromisso com a transparência exige também uma abordagem sistêmica à monitoração e à avaliação. É preciso compreender, a este respeito, que as avaliações aferem apenas o desempenho das crianças que estão nas escolas, pouco informando sobre a extensão e as causas de desistências. Outras ferramentas que precisam ser utilizadas na aferição do desempenho dos estudantes e do sistema incluem: estudos sobre desistência e repetência; mensuração dos insumos escolares e estimativas dos recursos mínimos que toda escola deve ter; observação sistemática dos processos escolares; e estudos do desempenho no mercado de trabalho dos que deixam as escolas.

Usos das Avaliações

O mais importante desafio para o futuro na América Latina será

assegurar que as avaliações sejam, de fato, usadas como instrumentos para melhorar a qualidade da educação. As avaliações podem influenciar vários elementos do sistema educacional, aí incluídos os programas de reforma educacional; a política educacional nacional; o currículo escolar; as decisões dos pais, estudantes e professores; as políticas educacionais locais e regionais; e a pedagogia e os programas no nível das escolas. Três áreas de particular importância na América Latina nas quais as avaliações podem ser aplicadas são o reforço do “alinhamento”, a destinação dos recursos para as escolas mais necessitadas e o encorajamento da responsabilidade e premiação dos progressos no nível da escola individual.

Reforço do “Alinhamento”

Acima de tudo, as avaliações podem ajudar os países da América Latina a melhor alinhar o currículo pretendido, o currículo real, os livros-textos, o conhecimento dos professores, a pedagogia de sala de aula e o aprendizado. Os ajustamentos resultantes farão que seja possível estabelecer metas mais altas, porém factíveis, aprimorar o

treinamento do professor e a pedagogia, e assegurar que o currículo revisado seja adotado na prática. Isto exigirá um grande esforço de análise, divulgação e treinamento. Com amplos recursos e *feedback*, as avaliações podem, então, tornar-se instrumentos diretos para a melhoria do ensino e do aprendizado. Até o presente, a Argentina foi quem mais avançou neste processo.

Destinação dos Recursos

Em vista da grande disparidade em resultado educacional através da América Latina, as avaliações devem ser usadas para canalizar recursos adicionais para as escolas mais necessitadas da região, especialmente as da área rural e das favelas urbanas. Esta atividade pode seguir as linhas mestras do programa das “900 escolas” do Chile, que é explicitamente voltado para a melhoria das escolas de pior desempenho.

Encorajamento de Melhores Desempenhos nas Escolas

As avaliações devem fornecer informações úteis às escolas, aos pais

e aos professores. Isto exigirá, entre outros elementos, a determinação do “valor agregado” da escolarização, através da identificação das escolas que apresentam melhor desempenho, considerado o *background* socioeconômico de seus alunos, como é feito na França, ou através da mensuração do progresso da pontuação das escolas ao longo do tempo. Estas abordagens levam ao reconhecimento e à premiação das escolas “eficazes”, e pressionam as demais escolas a copiar as mesmas características. Esta abordagem também evita preconceitos contra as escolas de zonas pobres. Até o momento, o Chile é quem mais fez para identificar e encorajar tais escolas.

Capacitação e Competência Técnica

Fortalecimento das Instituições Independentes

Na maioria dos países da América Latina, os governos centrais teriam grande dificuldade em implementar diretamente um programa de avaliação de longo prazo, como é feito na França. Entre os obstáculos que tal sistema teria de transpor, estão os potenciais conflitos de interesse, os

salários inadequados e os procedimentos inflexíveis de contratação e administração. Uma estratégia mais adequada à região seria promover agências de testes competentes fora do governo, mantendo, porém, uma pequena agência governamental com conhecimento e grande eficiência para supervisionar suas atividades. A Colômbia e o Brasil parecem ter desenvolvido parcerias deste tipo entre os setores público e o privado.

Garantia de Especialização Técnica Adequada e Apoio à Pesquisa

Para terem sucesso na aferição do que pretendem medir, as avaliações exigem um alto grau de especialização. Caso contrário, é bastante provável que não atinjam seu objetivo: podem não discriminar adequadamente os estudantes; podem prejudicar a possibilidade de comparação de um ano com o outro; as escolas e regiões podem trapacear; a estrutura da amostra pode estar incorreta, inutilizando toda a avaliação; e técnicas de análise mais sofisticadas podem revelar que supostas relações causais são espúrias. Muitos destes problemas já foram detectados em avaliações nacionais. É essencial, portanto, treinar e remunerar

adequadamente especialistas em teste de currículos, em metodologia de amostragem e pesquisa e em técnicas de análise. Os jordanianos concordariam com a afirmação de que é também importante consultar as mais renomadas e respeitadas autoridades mundiais em avaliação, já que não existe uma metodologia “regional” de avaliação.

Os governos e fundações precisam também fortalecer os centros de excelência de graduação em didática e pesquisa da região. Há, particularmente, necessidade de apoiar pesquisa independente baseada nos dados das avaliações, uma rica fonte de informações sobre todos os aspectos da educação, correlacionando-os com estudantes, professores e escolas. Para obter melhores resultados, esta pesquisa deve basear-se em técnicas atualizadas de análise. Igualmente importante, pesquisadores e formuladores de política precisam acordar os objetivos da pesquisa e estar dispostos a divulgar os resultados, mesmo se forem ambíguos, ou negativos.

Novas Abordagens

Participação em Programas Internacionais de Avaliação

O TIMSS oferece aos países da América Latina uma enorme oportunidade de estabelecer padrões, alinhar currículos e livros-textos, promover os professores e conseguir avanços importantes no aprendizado. Toda a região pode se beneficiar da experiência e da metodologia do TIMSS.

A IEA está atualmente planejando repetir o TIMSS, em 1999. O Chile e o Brasil pretendem participar. Os países da América Latina podem também participar do novo programa de educação em cidadania da IEA, que terá particular importância em questões relacionadas com o fortalecimento da sociedade civil na região. Atualmente, Chile e Colômbia estão engajados com o estudo da educação para cidadania.

Alternativamente, os pesquisadores do TIMSS podem revisar os currículos em textos e guias de currículo, comparando-os com os currículos de outras regiões do mundo. Os pesquisadores do TIMSS podem, ainda, tentar determinar se os exames nacionais, e outros exames, estão em linha com o currículo

nacional pretendido. A metodologia do TIMSS já foi utilizada no exame de conteúdo de currículos, livros-textos e pedagogia, na Argentina, na Colômbia, no México e na República Dominicana. A equipe internacional do TIMSS e os pesquisadores regionais poderiam ser convidados a fornecer assistência técnica a qualquer uma das mencionadas atividades.

Aplicação Universal de Testes versus Aplicação por Amostragem

Antes de iniciarem testes universais, os países devem estar conscientes de que esta é uma empreitada onerosa e que requer administração atenta e capaz. É preciso, também, um minucioso sistema de *feedback* e assistência às escolas individuais. No Chile, a aplicação universal dos testes custa cerca de US\$5.00 por estudante, ou cerca de 2 por cento dos custos totais da unidade. Mesmo este pequeno valor pode ser difícil de justificar em países que enfrentam severas restrições fiscais e onde faltam fundos para material educativo.

Em comparação, as pesquisas por amostragem custam menos e podem, ainda assim, fornecer

informações adequadas para o estabelecimento de políticas nacionais e para identificação dos problemas de alinhamento entre currículo, livros-textos e ensino. Tais pesquisas, contudo, exigem cuidadosa amostragem e administração rigorosa. Sempre que possível, os países da América Latina devem também utilizar para fins de avaliação os exames de seleção existentes.

Teste de Desempenho

Em anos recentes, os Estados Unidos vêm incorporando cada vez mais perguntas abertas em seus testes de desempenho, enquanto que a Europa tem-se concentrado em melhorar a confiabilidade e a validade dos instrumentos de medição. Em contraste, a maior parte da América Latina ainda prefere o tipo de avaliação através da múltipla escolha. Os países da região deveriam, ao menos, explorar as mais recentes abordagens de aplicação de testes. Não poderão fazê-lo, contudo, a menos que as pessoas que irão pontuar os testes recebam treinamento completo sobre os procedimentos.

Exames de Competência Mínima

Os países da América Latina podem também desejar considerar exames de competência mínima para o nível secundário. Tais exames foram já desenvolvidos em partes dos Estados Unidos e na Costa Rica, como um meio de encorajar o aprendizado. Exames deste tipo seriam especialmente apropriados para os países maiores, descentralizados, como o Brasil.

Opções para Países Menores, Mais Pobres

Os países menores e mais pobres não têm, simplesmente, o conhecimento técnico ou os recursos para implantar uma avaliação sustentada. Como o sugere a experiência da Costa Rica, os especialistas técnicos locais podem ser insuficientes para dar uma cobertura adequada. Assim, os países mais pobres devem visar um pequeno número de testes (por exemplo, em matemática e linguagem apenas), em poucas séries. Outras alternativas são: operar em consórcio

(composto, digamos, pelos países da América Central), aderir a programas internacionais, e aproveitar-se de assistência técnica regional (do Chile e da Argentina, por exemplo).

Cooperação Regional

Para fortalecer o processo de coordenação regional, será necessário estreitar os laços com a IEA e com os principais centros mundiais de pesquisa e desenvolvimento nas áreas de testes e avaliações, e estabelecer ampla participação, inclusive de instituições não-governamentais, em foros e comitês regionais. São necessários esforços regionais também para: analisar os resultados das avaliações da UNESCO/Orealc, de modo a identificar os fatores associados a bom desempenho; entender as interações entre objetivos curriculares e aprendizado, através da aferição da “oportunidade de aprendizado”; e, mais importante, identificar e apoiar centros regionais de excelência em testes, aferição e pesquisa educacional.

BIBLIOGRAFIA

- Alvarez, Benjamin, e Mónica Ruiz-Casares, eds. 1997. *Assessment and Education Reform: Policy Options*. 1997. Washington D.C.: Academy for Educational Development.
- Capper, Joanne. 1996. "Testing to Learn, Learning to Test". Newark, Del.: International Reading Association.
- Esquivel, Juan Manuel. 1996. "Medición de logros del aprendizaje y el empleo de los resultados en Costa Rica". PREAL, Seminário Avaliação e Determinação de Padrões na Educação Latino Americana. Helena Maria Bousquet Bomeny (org.), *Avaliação e Determinação de Padrões na Educação Latino-americana, Realidades & Desafios*, Rio de Janeiro, 1997, Editora Fundação Getúlio Vargas.
- Horn, Robin, Laurence Wolff, e Eduardo Velez. 1991. *Developing Educational Assessment Systems in Latin America: A Review of Issues and Recent Experience*. Washington, D.C., Banco Mundial.
- Ministério da Educação e Cultura, Brasil. 1996. "Seminário Internacional de Avaliação da Educação", outubro de 1995, Rio de Janeiro.
- de Moura Castro, Cláudio et. al. 1984. "Determinantes de la Educación en América Latina: Acceso, Desempeño y Equidad" – un Estudio ECIEL. (Programa de Estudos Conjuntos para Integração Econômica Latino-americana). Rio de Janeiro: Fundação Getúlio Vargas.
- Murphy, Paul, Vincent Greaney, Marlaine E. Lockheed, e Carlos Rojas, eds. 1996. *National Assessments – Testing the System*. Washington D.C., Banco Mundial, Instituto de Desenvolvimento Econômico.
- National Research Council. 1995. *National Science Education Standards*. Washington, D.C., National Academy Press.
- Rodríguez Schuller, Cristina. 1996. "Sistema de Medición de la

- Calidad de la Educación: características y uso de los resultados de la evaluación para mejorar la calidad (Chile)". PREAL, Seminário Avaliação e Determinação de Padrões na Educação Latino Americana. Helena Maria Bousquet Bomeny (org.), *Avaliação e Determinação de Padrões na Educação Latino-americana, Realidades & Desafios*, Rio de Janeiro, 1997, Editora Fundação Getúlio Vargas.
- Schmelkes, Silvia, 1996. "Propuesta para el diseño y desarrollo del sistema nacional de evaluación integral de la educación básica en México". PREAL, Seminário Avaliação e Determinação de Padrões na Educação Latino Americana. Helena Maria Bousquet Bomeny (org.), *Avaliação e Determinação de Padrões na Educação Latino-americana, Realidades & Desafios*, Rio de Janeiro, 1997, Editora Fundação Getúlio Vargas.
- Schmidt, William A., Curtis McKnight, e Senta A. Raizen. 1997. *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht: Kluwer Academic.
- Schmidt, William A. et. al. 1997. *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science*. Dordrecht: Kluwer Academic.
- U.S. National Research Center for the Third International Mathematics and Science Study. 1997. *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht: Kluwer Academic.
- Valdés, Eduardo Aldana. 1996. "Medición del logro educativo en Colombia". PREAL, Seminário Avaliação e Determinação de Padrões na Educação Latino Americana. Helena Maria Bousquet Bomeny (org.), *Avaliação e Determinação de Padrões na Educação Latino-americana, Realidades & Desafios*, Rio de Janeiro, 1997, Editora Fundação Getúlio Vargas.

Valverde, Gilbert A. 1996. "Las políticas educacionales nacionales y la participación en estudios internacionales". PREAL, Seminario Avaliação e Determinação de Padrões na Educação Latino Americana. Helena Maria Bousquet Bomeny (org.), *Avaliação e Determinação de Padrões na Educação Latino-americana, Realidades & Desafios*, Rio de Janeiro, 1997, Editora Fundação Getúlio Vargas.

Vianna, Heraldo Marelim. 1996. "Alguns programas de avaliação educacional no estado de São Paulo". PREAL, Seminario Avaliação e Determinação de Padrões na Educação Latino Americana. Helena Maria Bousquet Bomeny (org.), *Avaliação e Determinação de Padrões na Educação Latino-americana, Realidades & Desafios*, Rio de Janeiro, 1997, Editora Fundação Getúlio Vargas.

APÊNDICE: UMA PASSADA DE OLHOS SOBRE OS SISTEMAS DE AVALIAÇÃO EM SEIS PAÍSES

ARGENTINA

Data da primeira avaliação: 1993.

Data do mais recente programa de avaliação: 1995.

Séries avaliadas (programa mais recente): 2^a, 3^a, 5^a, e 7^a do primário.

Matérias avaliadas: matemática e leitura e, na 7^a série, também ciências naturais e sociais.

Amostra ou universo: universo dos estudantes.

Resultados importantes: os itens foram selecionados com base nos níveis mínimos esperados, mas os alunos responderam apenas 50 por cento das questões corretamente; fatores escolares foram responsáveis por 40 por cento da variação associada às notas de matemática.

Tipos de pesquisas realizadas: correlações e tabulações cruzadas simples.

Informações disponíveis sobre os custos: ---

Elementos técnicos de interesse: múltipla escolha e perguntas abertas.

Estrutura institucional: administrada pelo Ministério da Educação, mas sub-contratada com várias instituições privadas.

Feedback: programa compreensivo fornecendo informações sobre desempenho das escolas, perfis dos alunos e professores, modelos de gerência associados aos desempenhos; identifica conteúdo com maior dificuldade e inclui recomendações para melhoramento pedagógico.

Impacto na política: manuais metodológicos para aprimoramento do aprendizado, bem como para treinamento no serviço baseado nos resultados dos testes e distribuídos a todas as escolas; assistência técnica em todo o país.

Questões políticas e outras questões: ---

Relação com programas internacionais de avaliação: participou na análise de currículo do TIMSS e realizou os estudos tarde demais para serem incluídos no relatório; prestou assistência técnica à Bolívia e ao Paraguai.

BRASIL

Data da primeira avaliação: 1993.

Data do mais recente programa de avaliação: 1995.

Séries avaliadas (programa mais recente): 4^a e 8^a do primário e 2^a e 3^a do secundário (além disso, o Brasil iniciou um programa de avaliação de todos os estudantes universitários).

Matérias avaliadas: matemática e leitura.

Amostra ou universo: amostra de 90.499 estudantes em 2.289 escolas públicas e 511 escolas privadas.

Resultados importantes: os alunos, de modo geral, demonstraram níveis de aprendizado muito abaixo dos definidos por professores, pesquisadores e especialistas; os pontos mais altos foram observados no sul do país, nas capitais, nas escolas urbanas e privadas, e em escolas diurnas de tempo integral; os alunos mais novos, brancos e asiáticos e os estudantes com professores mais bem treinados tiveram melhor pontuação.

Tipos de pesquisas realizadas: correlações e tabulações cruzadas simples.

Informações disponíveis sobre os custos: ---

Elementos técnicos de interesse : múltipla escolha.

Estrutura institucional: administrada pelo Ministério da Educação, mas sub-contratada com várias instituições privadas.

Feedback: relatórios postos à disposição da imprensa, mas não utilizados pelas escolas individuais.

Impacto na política: nenhuma decisão tomada até o momento com base no programa de avaliação.

Questões políticas e outras questões: o uso político dos resultados da avaliação, como concentração nas escolas com desempenho fraco, não é considerado adequadamente. São Paulo, Minas Gerais e Paraná começaram, recentemente, programas de avaliação.

Relação com programas internacionais de avaliação: participou no estudo da IAEA de matemática e leitura em 1993 e planeja participar no TIMSS-R.

CHILE

Data da primeira avaliação: 1981.

Data do mais recente programa de avaliação: 1991-96.

Séries avaliadas (programa mais recente): 4^a e 8^a.

Matérias avaliadas: leitura e aritmética nas 4^a e 8^a séries; amostra de 10 por cento em ciências naturais, história e geografia.

Amostra ou universo: de modo geral universal, exceto quanto ao observado acima.

Resultados importantes: os resultados continuam a demonstrar fraco desempenho nas escolas públicas municipais e nas escolas rurais. Alta correlação entre status socioeconômico e desempenho; houve significativa melhora nas notas das escolas incluídas no programa das “900 escolas”, que se concentra nas escolas de pior desempenho.

Tipos de pesquisas realizadas: correlações e tabulações cruzadas simples; há trabalho analítico mais sofisticado sendo conduzido atualmente.

Informações disponíveis sobre os custos: US\$5,00 por estudante avaliado.

Elementos técnicos de interesse: múltipla escolha até o momento; considerando a possibilidade de introduzir mais questões de desempenho, abertas, e ensaios; método estatístico sendo revisado para assegurar maior confiabilidade de um ano para outro.

Estrutura institucional: foi desenvolvida capacidade técnica no Ministério da Educação, em vez de utilizar agência autônoma.

Feedback: os relatórios estão constantemente sendo simplificados e adaptados para platéias específicas, especialmente no nível das escolas; diretores de escolas, professores e pais usam os resultados do teste; alguns pais usam os resultados do teste para escolher a escola; os elaboradores do currículo nacional estão utilizando os resultados para revisar o currículo.

Impacto na política: os elaboradores de currículo estão enfatizando o domínio das matérias nas áreas mais fracas; programas pedagógicos direcionados para as escolas de desempenho mais fraco (programa “900 escolas”) continuam sendo desenvolvidos.

Questões políticas e outras questões: o programa tornou-se tão institucionalizado que “assuntos secundários” estão ficando mais importantes: por exemplo, os professores podem estar se concentrando demais na “preparação para os testes” e as escolas podem estar distorcendo as informações sobre o nível de pobreza de seus alunos.

Relação com programas internacionais de avaliação: nenhuma, diretamente, até o momento, mas planeja participar no TIMSS-R.

COLÔMBIA

Data da primeira avaliação: 1990.

Data do mais recente programa de avaliação: 1994.

Séries avaliadas (programa mais recente): inicialmente as 3^a e 5^a séries, e posteriormente as 7^a e 9^a.

Matérias avaliadas: inicialmente, matemática e leitura; subseqüentemente também ciências naturais e sociais.

Amostra ou universo: amostra apenas.

Resultados importantes: os níveis cognitivos dos estudantes da 3^a e 5^a séries são baixos, comparados com o esperado, especialmente na área de capacidade crítica e de reflexão; os pontos mais elevados vieram das escolas urbanas e privadas, entre os que cursaram o pré-escolar, não repetiram, tinham livros em casa e pais com níveis relativamente mais elevados de educação; os alunos de turmas com professores de melhor nível de educação, com livros-textos disponíveis, professoras mulheres, *escuela nueva* e escolas integrais tiveram melhor desempenho.

Tipos de pesquisas realizadas: correlações e tabulações cruzadas simples, e algumas análises multi-variadas.

Informações disponíveis sobre os custos: ---

Elementos técnicos de interesse : múltipla escolha e perguntas abertas.

Estrutura institucional: *joint venture* entre o Ministério da Educação, o Instituto SER, a Universidade Nacional de Professores, e o Centro de Estudos Sociais.

Feedback: relatórios de pesquisa e resumos disponíveis, mas sem utilidade para a maioria das escolas ou municípios devido à base em amostragem.

Impacto na política: A lei de educação de 1994 incorpora resultados das avaliações, como o enfoque no pré-escolar, na educação dos professores e em livros-textos

Questões políticas e outras questões: considera que a extensão para cobrir o universo dos alunos seria muito onerosa e pesaria demais sobre a capacidade institucional, mas irá ampliar a amostra para fornecer mais informações às escolas e regiões avaliadas.

Relação com programas internacionais de avaliação: participação integral no TIMSS, com divulgação dos resultados; planeja participar na pesquisa de educação sobre cidadania da IEA.

COSTA RICA

Data da primeira avaliação: 1986.

Data do mais recente programa de avaliação: 1993-97.

Séries avaliadas (programa mais recente): 1^a, 3^a, 6^a, 9^a e 12^a.

Matérias avaliadas: 1^a série, preparo para o aprendizado; 3^a, 9^a e 12^a séries, todas as matérias; feita amostragem também de capacidades cognitivas e físicas em diversas séries.

Amostra ou universo: principalmente universo, exceto em capacidade cognitiva e física e em preparo para aprendizado.

Resultados importantes: ainda não disponíveis para o teste atual; os resultados anteriores revelaram que o desempenho dos alunos estava muito abaixo do esperado, em comparação com os objetivos do currículo nacional; os estudantes em escolas privadas, urbanas e acadêmicas tiveram melhor desempenho. Os estudantes apresentaram desempenho progressivamente pior, em comparação com os objetivos do currículo, à medida que aumentava seu tempo na escola.

Tipos de pesquisas realizadas: tabulações cruzadas simples.

Informações disponíveis sobre os custos: US\$1,50 – US\$ 7,50 por aluno avaliado em 1991.

Elementos técnicos de interesse : múltipla escolha, principalmente.

Estrutura institucional: agora, um empreendimento cooperativo entre o Ministério da Educação e o IIMEC da Universidade da Costa Rica.

Feedback: os primeiros programas de avaliação foram pouco divulgados; no novo programa, os relatórios serão elaborados visando múltiplas platéias.

Impacto na política: re-introdução, em 1988, da avaliação nacional ou regional, com peso grande nas notas finais dos alunos; usado como justificativa para aumento do financiamento da educação.

Questões políticas e outras questões: abordagem intermitente, de arrancos e paradas, dependendo da atitude dos sucessivos ministros da educação.

Relação com programas internacionais de avaliação: nenhuma, diretamente.

MÉXICO

Data da primeira avaliação: 1976.

Data do mais recente programa de avaliação: 1994.

Séries avaliadas (programa mais recente): todas as séries dos ensinos básico e secundário, bem como todos os professores.

Matérias avaliadas: matemática e leitura nas escolas; os professores foram avaliados na matéria de conteúdo e em conhecimento pedagógico.

Amostra ou universo: universal em 1994, amostra anteriormente.

Resultados importantes: à medida que os estudantes avançam na escola, apresentam desempenho progressivamente pior que o esperado; as pontuações mais altas foram alcançadas por crianças que haviam freqüentado o pré-escolar, que não estavam repetindo o ano, que freqüentavam escolas urbanas ou privadas, que tinham pais com melhor educação e que viviam em zonas urbanas. As crianças de escolas indígenas e comunitárias com instalações inadequadas e professores menos bem treinados tiveram o pior desempenho.

Tipos de pesquisas realizadas: correlações e tabulações cruzadas simples.

Informações disponíveis sobre os custos: US\$ 35,00 por estudante nas amostras selecionadas em 1990.

Elementos técnicos de interesse : múltipla escolha.

Estrutura institucional: administrada pela *Secretaría de Educación Pública* do Ministério da Educação.

Feedback: relatórios colocados à disposição do público, mas de difícil leitura; até 1990 os resultados da avaliação eram considerados confidenciais.

Impacto na política: até recentemente, nenhum impacto; as avaliações de professores estão vinculadas a pacto com o sindicato nacional para melhoria da qualidade da educação básica.

Questões políticas e outras questões: disseminação e utilização inadequada no nível das escolas.

Relação com programas internacionais de avaliação: participação integral no TIMSS, mas solicitou, no último momento, que os resultados não fossem informados.