

*PREAL*

*Programa de Promoção da Reforma Educativa na América Latina e Caribe*

*Partnership for Educational Revitalization in the Americas*

**Nº 22**

**Como os Sistemas Nacionais de  
Avaliação Educativa da América  
Latina Apresentam Seus  
Resultados? (Versão Resumida)\***

*Pedro Ravela*

Maio de 2003

Revisão Técnica: Helena Maria B. Bomeny e Raquel B. Emerique  
Tradução de Paulo M. Garchet, [paulo@garchet.com](mailto:paulo@garchet.com)

---

\* Este documento é uma versão resumida de um trabalho mais amplo publicado pelo PREAL em agosto de 2001. A versão completa pode ser consultada na página da Web do *Grupo de Trabajo de Estándares y Evaluación* do PREAL <http://www.grade.org.pe/gtee-preal/docs/Resultados.pdf>.

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.

Documentos editados pelo PREAL Brasil:

**Série PREAL DOCUMENTOS:**

- 01: *Educação Pré-Escolar na América Latina: o atual “Estado da Prática”*. Robert G. Myers.
- 02: *Reforma da Educação Vocacional e Técnica na América Latina*. David N. Wilson.
- 03: *A última Palavra em Padrões Nacionais para Educação*. Diane Ravitch.
- 04: *Salas de Aulas Eficazes, Escolas Eficazes: Uma Base de Pesquisa Para Reforma da Educação na América Latina*. Robert E. Slavin.
- 05: *As Atuais Reformas Educativas na América Latina: Quatro Atores, Três Lógicas e Oito Tensões*. Cecilia Braslavsky e Gustavo Cosse.
- 06: *Ensino e Aprendizagem: em Busca de Novos Caminhos*. Glória Calvo.
- 07: *Educação na América Latina: Problemas e Desafios*. Jeffrey M. Puryear.
- 08: *A Reforma da Educação Chilena: Contexto, Conteúdos, Implantação*. Cristian Cox.
- 09: *Descentralização Educacional: Questões e Desafios*. E. Mark Hanson.
- 10: *América Latina e o Desafio do Terceiro Milênio: Educação de Melhor Qualidade Com Menores Custos*. Inés Aguerro.
- 11: *Avaliações Educacionais na América Latina: Estágio Atual e Desafios Futuros*. Laurence Wolff.
- 12: *Um estudo sobre os princípios e a prática dos vales-educação*. Edwin G. West.
- 13: *Desenvolvimento de Índices Financeiros da Educação Internacionalmente Comparáveis: A Experiência da OECD<sup>NT</sup> e Suas Implicações Para o Mercosul*. Stephen M. Barro.
- 14: *Aspectos Políticos das Implantações de Reformas Educativas*. Javier Corrales.
- 15: *Reformas Educativas na América Latina. Balanço de Uma Década*. Marcela Gajardo.
- 16: *Educação: Cenários de Futuro. Novas Tecnologias e Sociedade da Informação*. José Joaquim Brunner.
- 17: *Os Efeitos da Descentralização do Sistema Educacional Sobre a Qualidade da Educação na América Latina*. Donald R. Winkler e Alec Ian Gershberg.
- 18: *As Escolas Aceleradas: Uma Década de Evolução*. Henry M. Levin.
- 19: *Sindicalismo Docente e Reforma Educativa na América Latina na Década de 1990*. Guillermina Tiramonti.
- 20: *Os Próximos Passos: Como Avançar na Avaliação de Aprendizagens na América Latina?* Pedro Ravela (ed.), Richard Wolfe, Gilbert Valverde e Juan Manuel Esquivel
- 21: *Autonomia Escolar na Nicarágua: Restabelecendo o Contrato Social*. Gustavo Arcia e Humberto Belli
- 22: *Como os Sistemas Nacionais de Avaliação Educativa da América Latina Apresentam Seus Resultados?* Pedro Ravela.

**Série PREAL DEBATES**

- 01: *A Educação Para o Século XXI: o Desafio da Qualidade e da Equidade*. Maria Helena Guimarães de Castro.
- 02: *As Diretrizes Curriculares Nacionais Para o Ensino Médio*. Guiomar Namó de Mello.
- 03: *Obstáculos à Reforma Educacional no Brasil*. David N. Plank, José Amaral Sobrinho e Antônio Carlos da Ressurreição Xavier.
- 04: *A Nova Lógica das Lideranças Empresariais Sobre a Educação no Brasil*. Maria Teresa Cavalcanti de Oliveira.
- 05: *Contexto Socioeconômico e Descentralização Financeira no Ensino Fundamental em Minas Gerais*. Laura da Veiga, Bruno Lazzrotti Diniz Costa e Fátima Beatriz C. Teixeira P. Fortes.
- 06: *Desempenho Escolar e Desigualdades Sociais: Resultados Preliminares de Pesquisa*. Maria Lígia de Oliveira Barbosa.
- 07: *O Desenvolvimento da Ação Sindical do Ensino Privado Brasileiro*. Marcos Marques de Oliveira.
- 08: *Financiamento da Educação na América Latina: Lições da Experiência*. Alberto de Mello e Souza.
- 09: *Recursos Públicos em Educação*. Nicholas Davies.
- 10: *O Direito de Aprender*. Rudá Ricci.
- 11: *Seleção e Partilha: Desigualdades Sociais, Mérito e Excelência na Universidade*. Gláucia V. Bôas
- Preal Debates Especial: *Capital Social e Cultura: As Chaves Esquecidas do Desenvolvimento*. Bernardo Kliksberg

Peça seu exemplar:

PREAL Brasil. CPDOC – Fundação Getúlio Vargas.  
Praia de Botafogo, 190 sala 1410  
CEP 22253-900 – Rio de Janeiro - RJ  
Telefones: 2559-5678 FAX: 2551-2649  
E-mail: [bomeny@fgv.br](mailto:bomeny@fgv.br)  
[pronko@fgv.br](mailto:pronko@fgv.br)  
Home page: [www.cpdoc.fgv.br](http://www.cpdoc.fgv.br)

## ÍNDICE

### **Introdução/5**

#### **Capítulo I. O Que os Relatórios Nacionais se Propõem Realizar?/ 6**

- I.1. Os sistemas de avaliação como sistemas de informação ao público em geral/10
- I.2. Os sistemas de avaliação como sistemas de informação dirigidos a públicos específicos/11
- I.3. Usos dos resultados com conseqüências “fortes”/12
- I.4. O propósito central da divulgação dos resultados: como ”conseqüências fortes” ou como “avaliação formativa”/13

#### **Capítulo II. Percentuais, Médias, Notas e Pontuações: as Métricas Empregadas Para Informar os Resultados/16**

- II.1. Percentuais médios de respostas corretas e/ou médias brutas de pontuações?/17
- II.2. Percentuais de alunos que dominam determinados objetivos ou que atingem níveis pré-determinados de desempenho/19
- II.3. Pontuações padronizadas/21
- II.4. Teoria de Resposta ao Item/22
- II.5. Qualificações/26
- II.6. Resultados descritivos ou padrões de resultados: o que é que o público e os docentes desejam saber?/27

#### **Capítulo III. Comparando o Incomparável: a Parábola das Estradas/28**

- III.1. Modos típicos de desagregar a informação nos relatórios nacionais/28
- III.2. A parábola das estradas/29
- III.3. Alguns exemplos de conclusões impróprias/30
- III.4. A consideração do contexto sociocultural/35
- III.5. Os dilemas que a relação entre aprendizados e contextos sociais colocam na divulgação dos resultados/36

#### **Capítulo IV. A Utilidade Pedagógica dos Relatórios de Resultados/37**

- IV.1. Materiais de orientação didática elaborados a partir dos resultados/38
- IV.2. Apresentação de resultados por estabelecimento/41
- IV.3. Avaliações por amostragem com aplicação e correção autônomas das provas no universo de escolas/43
- IV.4. Seminários de divulgação dos resultados e programas de capacitação em serviço/45
- IV.5. Como usar as avaliações para melhorar as práticas de ensino?/45

#### **Capítulo V. Os Relatórios de Resultados Como Contribuição Para a Formulação de Políticas Educativas/46**

- V.1. Análise, interpretação e conceituação insuficientes/47
- V.2. Os “fatores escolares” associados aos aprendizados e às políticas educativas/ 49
- V.3. Três deficiências de caráter técnico comuns na região/ 54

### **Dez Sugestões Para Pensar o Futuro/54**

### **Bibliografia/57**

## Introdução

Em 1999 o *Grupo de Trabajo sobre Estándares y Evaluación – GTEE* (Grupo de Trabalho sobre Padrões e Avaliação) do PREAL produziu um primeiro documento de análise sobre a situação dos sistemas nacionais de avaliação de aprendizados na América Latina, documento este no qual propusemos algumas medidas essenciais para se avançar no desenvolvimento de tais sistemas<sup>1</sup>. Esse trabalho foi elaborado no contexto de um seminário de trabalho realizado em Lima, em agosto de 1999, recorrendo-se basicamente ao conhecimento direto dos participantes sobre a experiência de diversos países em matéria de avaliação de aprendizados em nível nacional.

O presente documento constitui, de certo modo, um aprofundamento daquela análise a partir do estudo sistemático de um amplo espectro de Relatórios Nacionais da região no período de 1995-2000. Foram analisados quase 70 relatórios, no total<sup>2</sup>.

A análise do estudo concentrou-se nas seguintes perguntas:

- Que tipos de informações são produzidas e divulgadas pelos sistemas de avaliação?
- Como essas informações são apresentadas?
- A quem se dirigem?
- Com que propósitos?
- Como se espera que as informações sejam utilizadas?

---

<sup>1</sup> RAVELA, P. (editor), WOLFE, R., VALVERDE, G. e ESQUIVEL, J.M., (2000). “Los Próximos Pasos: ¿hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?” PREAL/GRADE: Lima.

<sup>2</sup> Por questão de espaço não se inclui nesta “versão resumida” a relação completa dos Relatórios analisados. Ela pode ser consultada na versão completa, na Página do GRADE na Web.

- As informações servem efetivamente a tais propósitos?

- Como melhorar nossos modos de informar os resultados?

O documento está organizado em cinco capítulos.

No primeiro, analisam-se os objetivos – das avaliações e dos relatórios – que são explicitamente assinalados nos Relatórios Nacionais. Discute-se a principal alternativa em termos da política de avaliação, que consiste em decidir se as provas devem ter algum tipo de conseqüências diretas, ou “fortes”, para as escolas ou se devem ser, principalmente, de caráter formativo.<sup>3</sup>

O segundo capítulo apresenta uma visão sistemática das “métricas” – ou tipos de dados como percentuais de respostas corretas, pontuações ou percentuais de alunos que dominam certos objetivos – empregadas na região para informar os resultados. Muitas vezes, o leitor não especializado terá dificuldade em distinguir os significados dos diferentes tipos. Assim, ilustra-se e explica-se o significado de cada um e discute-se que tipo de enfoque é mais adequado para atender à pergunta principal que, em princípio, os sistemas de avaliação devem ajudar a responder: nossos alunos aprenderam o que se supunha que deveriam ter aprendido?

No terceiro capítulo, analisa-se uma das principais deficiências dos relatórios de resultados: a falta de atenção adequada às diferenças socioculturais existentes entre as populações atendidas, ou pelos distintos setores do sistema educativo, ou pelos diversos tipos de escolas. Os alunos chegam à escola com marcantes diferenças quanto ao domínio da linguagem padrão e suas capacidades simbólicas e, ao longo de sua permanência na escola têm diferentes

---

<sup>3</sup> O que na literatura anglo-saxônica é conhecido como *high-stakes* contra *low-stakes*.

condições de vida e de níveis de apoio em casa para seus deveres, fatores que favorecem ou dificultam o trabalho da escola. Assim, comunicar os resultados ignorando essas diferenças pode levar a conclusões grosseiramente impróprias, a principal das quais talvez seja que as escolas privadas seriam melhores que as públicas.

Destacamos dois dos principais propósitos explícitos dos relatórios de resultados: dar aos docentes informações relevantes para melhoria do ensino; e municiar as autoridades de informações úteis para a tomada de decisões e a formulação de políticas. O quarto capítulo analisa quatro tipos de estratégias desenvolvidas para o primeiro destes propósitos, enquanto o quinto é dedicado ao segundo. Em ambos os casos analisam-se os pontos fortes e fracos das informações contidas nos relatórios nacionais em relação a sua potencial utilidade para os destinatários definidos.

Cada ser humano é produto de sua circunstância, do contexto em que vive e, inevitavelmente, sua visão do mundo será influenciada por essa circunstância. Neste sentido, desejo assumir explicitamente que a perspectiva em que este documento foi escrito é duplamente parcial. De um lado porque vejo os relatórios nacionais contra o pano de fundo da experiência que realizamos no Uruguai e das prioridades que assumimos para nosso sistema de avaliação. De outro, porque vejo a experiência de outros países com parcialidade, a partir da leitura de alguns relatórios. Assim, peço desculpas de antemão pelas omissões ou erros de perspectiva em que seguramente terei incorrido ao analisar os relatórios. Desejo manifestar, também, que o trabalho baseou-se, fundamentalmente, em relatórios publicados até o ano de 1999, e que desde então vários países mudaram seus respectivos enfoques, mudanças estas que não são contempladas neste documento.

De todo modo, quero enfatizar que as referências críticas a certos aspectos dos relatórios nacionais não pretendem desqualificar ou julgar o trabalho realizado pelas unidades aludidas, mas, simplesmente, ilustrar deficiências e dificuldades, boa parte

das quais compartilhamos todos os países da região e cuja superação devemos assumir como desafio coletivo. A intenção deste trabalho é contribuir para a reflexão e a acumulação de conhecimento na região sobre como melhorar nossos sistemas de avaliação de aprendizado e nossos modos de comunicar os resultados de forma que tenham um impacto real na melhoria do aprendizado das crianças, nas práticas de ensino dos professores e na tomada de decisões de política educativa.

Finalmente, desejo agradecer especialmente a Santiago Cueto, Richard Wolfe, Gilbert Valverde e Patricia Arregui por seus enriquecedores comentários à versão preliminar deste documento.

## Capítulo I.

### O que os relatórios nacionais se propõem realizar?

Um primeiro exame dos relatórios de resultados das avaliações nacionais consiste em verificar como eles próprios se apresentam e que propósitos afirmam perseguir, tanto os relatórios especificamente, como, de forma mais geral, os sistemas de avaliação de aprendizado.

Na literatura sobre avaliação das instituições dos sistemas educativos costuma-se fazer distinção entre:

- avaliações cujos resultados terão conseqüências diretas para os indivíduos ou instituições e, por conseguinte, implicam um certo nível de risco ou ameaça porque implicam algum tipo de sanção positiva ou negativa;
- avaliações cujos resultados têm função unicamente informativa e formativa, mas não prevêm conseqüências.

As primeiras são chamadas “high stakes” e as segundas “low-stakes” (MESSICK, 1999), expressões que poderiam ser traduzidas como “alto” e “baixo” riscos, respectivamente. A figura 1 apresenta esquematicamente uma tipologia dos modos

como os resultados das avaliações podem ser utilizados para promover a melhoria da

qualidade dos sistemas educativos.

**Figura 1**  
**Tipologia de usos possíveis das avaliações nacionais para melhorar o sistema educativo**

| BAIXO RISCO   | ALTO RISCO   |
|---|--|
| Informação ao público sobre o estado do sistema educativo através da apresentação periódica, ou de resultados gerais nacionais, ou em grandes desagregados tipo urbano/rural, ou por regiões ou estados.. | Exames de certificação ao final de determinado nível educativo ou provas de seleção dos alunos que desejam ingressar no nível superior, ou terciário.                                    |
| Contribuição para a definição das políticas educativas através do fornecimento de informações aos responsáveis pelas decisões nos níveis superiores e intermediários.                                     | Estabelecimento de incentivos econômicos para as escolas ou para os professores a partir dos resultados obtidos por seus alunos nas provas e outros indicadores                          |
| Contribuição para o melhoramento das escolas e o desenvolvimento profissional dos docentes através da devolução dos respectivos resultados às escolas e da produção de materiais de caráter didático.     | Divulgação dos resultados de cada estabelecimento na imprensa com o intuito de permitir que as famílias possam julgar a qualidade das escolas e escolher a que desejam para seus filhos. |

A análise dos objetivos e propósitos incluídos explicitamente nos relatórios nacionais permite constatar que **a maioria dos sistemas nacionais de avaliação desenvolvidos na América Latina durante a década de 1990 enfatiza finalidades de caráter formativo (“low stake” ou “baixo risco”), sem conseqüências diretas para as escolas.**

De tal análise depreende-se que há dois grandes tipos básicos de papéis explícitos para os sistemas nacionais de avaliação:

- constituírem-se em sistemas de informação ao público sobre o estado da educação nacional;
- constituírem-se em provedores de informações para a tomada de decisões por parte de atores específicos – autoridades, supervisores, diretores, professores, famílias.

No quadro 1 estão consignadas, a título de ilustração, algumas das definições tiradas dos relatórios e que representam o tom que caracteriza a grande maioria deles.

**Quadro 1: Finalidades e propósitos dos sistemas nacionais de avaliação  
BRASIL - Sistema de Avaliação da Educação Básica (SAEB)**

*“Os dados do SAEB permitem:*

- *inferir os conhecimentos e habilidades dos alunos nas áreas e disciplinas investigadas através da aplicação de testes, não com a intenção de “avaliar” o aluno, mas com o objetivo de identificar o que o aluno é capaz de fazer nos distintos momentos de sua trajetória escolar (proficiência), com a finalidade de ponderar a qualidade e a equidade do ensino ministrado;*
- *analisar a eficácia dos currículos em função dos resultados de aprendizagem dos alunos;*
- *verificar os fatores contextuais e escolares que incidem na qualidade do ensino básico no país, utilizando questionários que recolhem dados sobre as características da infra-estrutura e dos recursos de que dispõe a unidade escolar, sobre o perfil do diretor e os mecanismos de gestão escolar, o perfil do professor e sua prática pedagógica, e sobre as características socioculturais e os hábitos de estudo do aluno.*

*A análise dos resultados dos levantamentos do SAEB permite acompanhar a evolução do desempenho dos alunos e dos diversos fatores que influenciam a qualidade e a eficácia do ensino ministrado no sistema educativo, possibilitando a definição de ações dirigidas à correção das distorções identificadas e ao aperfeiçoamento das práticas e dos resultados apresentados pelas escolas e pelo sistema de ensino brasileiros. Essas informações são utilizadas pelos gestores e administradores da educação, pesquisadores e professores”.*

MEC-INEP. Bases para um Ensino de Qualidade. Brasil, 1999.

**CHILE - Sistema de Medición de Calidad de la Educación (SIMCE)**

*“O SIMCE mede para melhorar a qualidade da educação. Com este propósito, fornece informações sobre os resultados de alguns aspectos do processo educativo, tanto no nível do estabelecimento, como nos níveis provincial, regional e nacional, para suas análises e interpretações e para a formulação de planos de ações corretivas eficientes, concretas e avaliáveis (...) A análise conjunta, unida à experiência profissional e ao conhecimento que os professores têm de sua Unidade Educativa, fará que os resultados obtidos sejam uma valiosa fonte de informações para formulação de ações corretivas mais eficazes”.*

Ministerio de Educación, República de Chile. Boletín Informativo N°5 para Profesores. Educación Básica. Análisis e Interpretación de Resultados 1996.

**EQUADOR - Sistema Nacional de Medición de Logros Académicos (APRENDO)**

*“O relatório técnico “APRENDO 97” pode servir a uma variedade de usuários, segundo seus respectivos interesses e atividades: os líderes e gestores do sistema educativo podem empregá-lo como fonte de informações confiáveis para tomada de decisões em nível de gestão financeira, administrativa e curricular; os professores e professoras podem utilizar a informação com insumo nos processos de planejamento curricular no nível micro e de avaliação do aprendizado; os pais de família e a sociedade civil em geral podem servir-se do relatório para melhor conhecer e compreender a realidade do sistema educativo equatoriano; finalmente, a comunidade acadêmica pode encontrar no relatório técnico hipóteses e dados valiosos para empreender projetos de pesquisa aplicada no campo da educação”.*

Ministerio de Educación y Cultura del Ecuador – EB/PRODEC, 1998.

## **I. 1. Os sistemas de avaliação como sistemas de informação ao público em geral**

A primeira grande intenção compartilhada por todos os sistemas nacionais de avaliação é informar à opinião pública até que ponto o sistema educativo vem atingindo certos objetivos em determinadas áreas do currículo e em certos graus ou níveis (geralmente aqueles que implicam o final de um ciclo de ensino).

Nesta perspectiva, as Unidades de Avaliação teriam uma função comparável à dos Institutos Nacionais de Estatística: assim como estes apresentam ao público, periodicamente, informações sobre o custo de vida (índice de preços ao consumidor) ou sobre o mercado de trabalho (índice de desemprego), **a função principal dos relatórios nacionais seria a de apresentar ao público informações periódicas sobre o aprendizado no sistema educativo, assegurando sua qualidade técnica.** O uso e a interpretação das informações seria responsabilidade primordial dos usuários, do mesmo modo que os Institutos de Estatística não se preocupam normalmente, nem em garantir que a população compreenda as informações que produzem, nem em fazer um trabalho “didático” de divulgação – trabalho este que geralmente fica a cargo da imprensa.

No caso da avaliação de aprendizados, é discutível que se possa adotar tal premissa. Em primeiro lugar, porque não existe acumulação suficiente de massa crítica a respeito das formas de avaliação e dos indicadores a serem usados. Ao contrário dos índices de população, emprego, evolução econômica, etc. – aos quais dedicou-se bastante trabalho acadêmico e em torno dos quais obtiveram-se consensos básicos quanto às formas de medição e interpretação e quanto a suas debilidades e limitações –, nas avaliações de aprendizado ainda se está longe de chegar a acordos semelhantes sobre os enfoques básicos, nos países da região. Há uma grande diversidade de modos de medir, informar e interpretar os dados e são necessários esforços de discussão e

acumulação técnica que deveriam incluir necessariamente a abertura dos instrumentos e procedimentos de avaliação ao escrutínio acadêmico.

Em segundo lugar, os indicadores de tipo demográfico e econômico passaram, em alguma medida, a fazer parte da cultura geral e têm um significado pelo menos genérico para os cidadãos medianamente informados<sup>4</sup>. Isto não ocorre com os indicadores de resultados da educação. Por isto, **é necessário realizar um trabalho sistemático de divulgação para que o público em geral possa atribuir significado aos dados apresentados pelos relatórios das avaliações.** Em especial, é imprescindível realizar tal trabalho com os meios de comunicação, com o intuito de criar uma certa cultura básica em torno do significado da avaliação de aprendizados, em nível do sistema e de seus resultados, de modo a evitar simplificações no tratamento das informações que chegam ao grande público.

É possível afirmar, então, que nos países da região foi dado um importante primeiro passo: a sociedade está sendo informada periodicamente sobre os resultados do aprendizado dos alunos, o que contribui para pôr a educação na agenda pública, levando a que a preocupação não se concentre mais exclusivamente no acesso ao sistema

---

<sup>4</sup> Ainda que em muitos casos, como o dos indicadores adequados para medir a pobreza, haja ainda considerável debate. Além disto, toda interpretação do estado de uma sociedade ou de uma economia a partir de um conjunto de índices é sempre matéria discutível. Os indicadores não nos apresentam a realidade, simplesmente nos apontam certos pontos de referência empíricos para que possamos interpretá-la. Do mesmo modo, nenhum sistema de indicadores educativos ou de resultados de avaliações pode, nem eliminar a necessidade do debate, nem pretender dar conta de maneira indiscutível e cabal do estado do sistema educativo. Não obstante esta ressalva, é de se desejar, sim, que os indicadores sejam pertinentes e que sua medição seja tecnicamente correta.

educativo, e passe a englobar o aprendizado que nele efetivamente se adquire. Simultaneamente, porém, é preciso assinalar que **as formas de comunicar os resultados desenvolvidos até o momento podem, em muitos casos, levar a sérios equívocos**, como se analisará detalhadamente mais adiante.

### **I. 2. Os sistemas de avaliação como sistemas de informação dirigidos a públicos específicos**

Além de informar os cidadãos de forma sistemática, **a maioria dos sistemas de avaliação assume como missão principal a de apresentar informações relevantes a três públicos específicos que, supostamente, irão usá-las para melhorar a qualidade da educação: as autoridades e os indivíduos que tomam decisões nos sistemas educativos, os diretores e docentes dos estabelecimentos educativos, e as famílias dos estudantes.**

O primeiro problema que surge é que alguns países pretendem informar esses distintos destinatários através de um relatório nacional único. Além disso, a maioria dos relatórios nacionais limita-se a descrever os dados incluídos nos quadros ou gráficos, com nenhuma ou escassas análise e interpretação, supondo que os mesmos sejam auto-explicativos. Em muitos relatórios, sequer são incluídas advertências sobre os usos válidos das informações e suas limitações.

Na figura 2 mostra-se o tipo de relatórios que cada país produziu. Chile, Equador, Argentina e Uruguai produzem relatórios dirigidos especificamente aos professores, com o intuito de dar feedback para análise de suas práticas pedagógicas através do estudo dos objetivos que apresentaram maior e menor resultados e de suas implicações pedagógico-didáticas. Em vários países, cada estabelecimento recebe os resultados de seus alunos, com a intenção de propiciar a análise e a tomada de decisões a partir das próprias

deficiências. Este aspecto tornará a ser visto, em maior profundidade, no capítulo IV.

Quanto às autoridades e aos encarregados de decisão nos níveis centrais, em muitos países foram produzidos estudos dos “fatores associados” aos resultados. Apesar disso, parece predominar uma visão excessivamente otimista – talvez um tanto ingênua – do quanto a pesquisa sobre “fatores associados” pode contribuir para a definição das políticas educativas. Na maioria dos relatórios presume-se uma relação direta e, de certo modo mecânica, entre a identificação dos “fatores” internos do sistema educativo associados aos resultados e a adoção de medidas para melhoria da educação. Um exemplo típico seria: *“Desenvolver ações diretas sobre os fatores controláveis pelo próprio sistema educativo que incidam no resultado das competências básicas dos alunos”*. Ainda que este propósito seja louvável, é preciso ter em mente que a relação entre os resultados de pesquisa e a tomada de decisões de política nunca é direta e imediata, sendo cortada por múltiplas mediações de tipo conceitual, social, político e administrativo. No capítulo V aprofundaremos a análise deste aspecto.

**Figura 2**

**Tipos de relatórios de divulgação de resultados produzidos nos países da região**

| <b>Países</b>    | Relatório geral de resultados | Relatórios específicos sobre fatores associados | Relatórios ou publicações dirigidas aos docentes | Relatórios ou publicações dirigidas às famílias | Relatórios de resultados por estabelecimento educativo (censitários) <sup>(*)</sup> |
|------------------|-------------------------------|---|--|---|---|
| <b>ARGENTINA</b> |                               |   |  |   |   |
| <b>BOLÍVIA</b>   |                               |   |  |   |   |
| <b>BRASIL</b>    |                               |   |  |   |   |
| <b>COSTA</b>     |                               |   |  |   |   |
| <b>CHILE</b>     |                               |   |  |   |   |
| <b>EQUADOR</b>   |                               |   |  |   |   |
| <b>PARAGUAI</b>  |                               |   |  |   |   |
| <b>PERU</b>      |                               |   |  |   |   |
| <b>URUGUAI</b>   |                               |   |  |   |   |

(\*) Mesmo no contexto de aplicações por amostragem, alguns países, como o Equador, entregam aos estabelecimentos que integraram a amostra seus respectivos resultados.

Quanto ao terceiro dos atores privilegiados nas declarações de propósito dos sistemas nacionais – as famílias dos estudantes – é preciso enfatizar que os relatórios nacionais dificilmente são compreensíveis para eles, mesmo quando se trata de uma família de nível cultural médio. O Chile parece ser o único país com uma estratégia de divulgação voltada especificamente para as famílias, estratégia esta que consiste na publicação dos resultados de todas as escolas, através da imprensa e da Internet, e na distribuição às famílias de folhetos explicativos com os resultados dos estabelecimentos onde seus filhos estudam.<sup>5</sup>

### I. 3. Usos dos resultados com conseqüências “fortes”

Ainda que os enunciados explícitos dos sistemas nacionais de avaliação enfatizem sua finalidade formativa e um uso de “baixo risco” dos resultados, há na região algumas experiências importantes de utilização dos

<sup>5</sup> Este último aparece na página do SIMCE na Web, mas não foi possível ter acesso a exemplares impressos do mesmo.

resultados das avaliações nacionais com conseqüências mais “fortes” que merecem ser identificados.<sup>6</sup>

O caso mais destacado é o do Chile onde, mesmo sem constar entre os propósitos explícitos dos relatórios nacionais, desde o início do SIMCE optou-se por utilizar os resultados das avaliações nacionais como mecanismo para aplicação de incentivos e sanções às escolas. Os resultados de cada estabelecimento são publicados na imprensa nacional – e, mais recentemente, na Internet – com o intuito, entre outros, de que os pais utilizem estas informações na hora de escolher a que escola encaminharão seus filhos, no contexto de uma política destinada a gerar um mercado competitivo entre os estabelecimentos educacionais. A suposição é que, desta forma, as escolas serão obrigadas a preocupar-se em melhorar seus resultados.

<sup>6</sup> Além disto, é preciso mencionar que, em vários países, os sistemas nacionais de avaliação coexistem com outros sistemas de provas de seleção, ou para admissão à universidade, ou para emissão de certificado de conclusão do curso secundário.

Além disto, desde 1996, os resultados das avaliações do SIMCE vêm sendo empregados para estabelecer incentivos econômicos para os professores, através do *Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales Subvencionados - SNED*. Os estabelecimentos educativos são classificados em função de um índice que inclui os seguintes fatores:

- *eficácia*: a média dos resultados do estabelecimento nas provas de Matemática e Espanhol aplicadas nos anos anteriores;
- *superação*: média das diferenças verificadas entre as pontuações dos estabelecimentos nas provas atuais em relação à aferição anterior;
- *iniciativa*: compreende diversos indicadores, tais como a realização de atividades formativas para os alunos, ou o desenvolvimento de um “projeto educativo institucional”;
- *melhoria das condições de trabalho*: constituída a partir a partir de uma categorização dos estabelecimentos realizada pelo sistema de inspeção;
- *igualdade de oportunidades*: compreende diversos indicadores, tais como a taxa de aprovação dos alunos, a taxa de retenção e a incorporação de alunos portadores de deficiências;
- *integração e participação de professores, pais e procuradores*: constrói-se a partir, ou de indicadores tais como a existência de um Conselho de Professores e de centros de pais de alunos que contem com espaços de participação garantidos no projeto institucional, ou da percepção que os pais têm do estabelecimento, entre outros.

Praticamente, dois terços do valor do índice dependem dos resultados das provas do SIMCE: o fator *eficácia* tem peso de 37% e *superação* 28% do valor final do índice SNED<sup>7</sup>.

Outro caso relevante é o do México. Durante a década de 1990, os esforços de avaliação de aprendizado no nível nacional concentram-se no programa *Carrera Magisterial*, programa que estabelece incentivos econômicos individuais para os professores, incentivos estes associados a cinco grandes fatores: antiguidade, grau acadêmico, preparação profissional, certificação em cursos de atualização e desempenho profissional. Este último fator, que tem o maior peso no índice (35%), é medido a partir da avaliação do professor por parte da equipe docente e do “aproveitamento escolar”, que é o resultado obtido pelos alunos de cada professor nas provas padronizadas de aprendizado (Comisión Nacional SEP-SNTE, 1994). A participação neste programa tem tido caráter voluntário, mas já abrange centenas de milhares de professores a cada ano. Durante muitos anos, as provas aplicadas a milhões de alunos foram utilizadas unicamente com a finalidade de outorgar estes incentivos, sem que seus resultados tivessem qualquer tipo de divulgação pública. Há muito pouco tempo, a *Secretaría de Educación Pública* começou a divulgar alguns resultados.

#### I. 4. O propósito central da divulgação dos resultados: “conseqüências fortes” ou “avaliação formativa”?

A discussão em torno de qual deve ser o objetivo principal dos sistemas de avaliação de resultados é de enorme transcendência e merece especial atenção na construção de visões sobre como avançar no desenvolvimento dos sistemas nacionais de avaliação de aprendizados da região.

Em alguns setores, principalmente entre políticos e economistas, parece existir crescente pressão por uma transformação gradual dos sistemas nacionais de avaliação em sistemas de prestação de contas públicas e de estabelecimento de incentivos ao estilo do que foi adotado no adotado no Chile. Recentemente, a Argentina começou a divulgar através da imprensa os resultados

<sup>7</sup> Para mais detalhes, ver Ministerio de Educación, República de Chile, 2000.

dos estabelecimentos individuais em suas avaliações de caráter censitário. Outros países – como o Uruguai e a Bolívia –, ao contrário enfatizaram exclusivamente a finalidade formativa das avaliações e mantêm os resultados por escola em caráter estritamente confidencial.

Em dois trabalhos recentes, referindo-se aos Estados Unidos – provavelmente o país que tem maiores experiências acumuladas, diversidade de enfoques e debates em torno deste tema – Linn & Gronlund (2000) assinalam que **o problema central de usar provas padronizadas no contexto de uma estratégia de “alto risco” ou “consequências fortes” (high stakes) – leia-se prestação de contas públicas ou de concessão de incentivos econômicos conforme os resultados – é que exacerbam a preocupação de diretores e professores em preparar para a prova e para o tipo de atividades que elas incluem**, vale dizer, a dedicar cada vez mais tempo de instrução a exercícios similares aos que serão apresentados nas provas, e aos temas e competências que serão avaliados. Se as provas tiverem uma forte carga de questões de múltipla escolha, isto acaba sendo altamente contraproducente, porque o resultado deixa de ser um indicador de um conjunto mais amplo de competências e passa a ser, estritamente, uma medida da capacidade dos alunos em responder às perguntas das provas. Não seria impossível melhorar as provas, fazê-las mais complexas e abrangentes, mas neste caso ficaria mais difícil garantir uma correção confiável das mesmas em grande escala já que, por definição, qualquer sistema de incentivos ou prestação pública de contas das escolas devem ter caráter censitário. Como assinalam Linn & Gronlund<sup>8</sup>:

*“Até que ponto esta crescente pressão [por fazer que as avaliações tenham “consequências fortes”] ajudou ou prejudicou a educação continua sendo motivo de*

*controvérsia. Os defensores de avaliações com*

*“consequências fortes” argumentam que as provas medem objetivos que são importantes para os alunos e nos quais é desejável que os professores concentrem sua atenção. Assinalam com orgulho os aumentos nas pontuações das provas que se verificaram nos programas estaduais e distritais de avaliação do final da década de 1980, início da de 1990.*

*“Os críticos da crescente ênfase nos resultados das avaliações afirmam que a importância excessiva dada às provas distorce a educação. Argumentam que objetivos educativos importantes são ignorados quando não contam pontos. Mais ainda, reclamam que o aumento das pontuações apresenta uma visão equivocada da realidade, porque os professores ensinam as questões específicas das provas, em vez de conteúdos mais gerais (...) Como as questões de uma prova constituem apenas uma amostra do domínio que interessa, a pontuação na prova e o grau de compreensão da matéria não são a mesma coisa. É preciso uma generalização, e é a generalização, não a prova, que importa. Quando se ensina para responder ao tipo de questões específicos das provas, a validade da inferência sobre o nível de aprendizado dos estudantes vê-se ameaçada”.*

A avaliação “sem consequências”, ou de “baixo risco” (*low stakes*) minimiza estes problemas, porque diretores e professores vêm-se menos ameaçados pela avaliação, com o que podem assumi-la com maior liberdade, como indicador de competências mais complexas, sem se verem obrigados a forçar artificialmente bons resultados nas provas.

Como contrapartida, é preciso assinalar que **a ausência de consequências faz que ninguém nas escolas fique obrigado a prestar contas dos resultados e, muitas vezes, nem sequer de inteirar-se deles ou de levá-los em consideração**. Neste sentido, é possível argumentar que a confidencialidade dos resultados por escola,

<sup>8</sup> LINN, R. & GRONLUND, N., 2000, pp. 8-9, em inglês no original.

de certo modo, exige de responsabilidade pública.

Outro dos problemas importantes derivados do fato de que as provas não tenham conseqüências diretas para os alunos é que isto pode afetar o grau de esforço que estes dedicam às tarefas que lhe são propostas. Isto é particularmente delicado no ensino médio, devido à idade dos alunos, e pode afetar a confiabilidade e a validade das medições.

Com relação ao uso dos resultados para gerar concorrência entre os estabelecimentos, CARNOY (2000) assinala que a evidência empírica do caso chileno indica que a criação de um mercado competitivo no setor educativo não teve impacto na melhora da qualidade do sistema, mas aumentou as desigualdades. Quando se controla o nível socioeconômico dos alunos de cada escola, observa-se que os colégios privados católicos, anteriores ao processo de privatização da década de 1980, têm resultados um tanto melhores que as escolas públicas, mas a um custo maior por aluno. Simultaneamente, os colégios privados “novos” – que “proliferaram como fungos” a partir da implantação dos “vouchers” (vale educação) nos anos oitenta – não apresentam resultados melhores que os das escolas públicas quando se controla o nível sociocultural dos alunos. São, isto sim, economicamente mais “eficientes” – têm um custo por aluno inferior – porque pagam menores salários a seus professores. A este respeito assinala:

*“Portanto, não há evidência, no Chile, de que o sistema nacional de “vouchers” (vales educação) que vem operando há mais de vinte anos tenha tido um efeito positivo sobre os resultados de rendimento dos milhões de estudantes que passaram pelo sistema durante este tempo. O plano pode ter economizado algum dinheiro de impostos na década de 1980, início da de 1990, mas, mesmo aí, somente porque as escolas privadas com fins lucrativos puderam pagar menos aos professores que as escolas públicas, e não porque houvesse mais eficiência no uso dos recursos. Por outro lado, há certa evidência de que a*

*privatização aumentou a estratificação, durante a década de 1980, na medida em que as escolas privadas atraíram os melhores estudantes – de nível social mais alto e mais talentosos – das escolas públicas e tenderam a localizar-se nos municípios de rendas mais altas... Estudos em outros países, como a Nova Zelândia, que têm sistemas similares de “vouchers”, também indicam que o uso desses vales-educação aumentam a estratificação entre os estudantes (Lauder & Hughes, 1999)”<sup>9</sup>*

Neste ponto, defrontamo-nos com outro dilema: ainda que a evidência empírica e o senso comum indiquem que a publicação da classificação das escolas em função de seus resultados tem como efeito principal uma maior segmentação sociocultural do sistema educativo, é preciso, ao mesmo tempo, reconhecer que, na medida em que se produzam informações sobre os resultados obtidos pelos alunos de determinado estabelecimento, suas famílias têm o direito de conhecê-las. Coloca-se, assim, uma contradição difícil de resolver e que os sistemas deverão assumir nos próximos anos: como conjugar o direito das famílias, de saber como seus filhos estão se desenvolvendo, com o cuidado que se deve dar ao tratamento das informações para evitar interpretações impróprias e a deslegitimação espúria das escolas que trabalham com a população mais carente?

A equação básica a ser resolvida é em que medida os objetivos *formativos*, de um lado, e de *estabelecimento de incentivos*, do outro, podem ser compatibilizados. O tema merece ser aprofundado através de uma pesquisa que analise o modo como as escolas utilizam os resultados das avaliações e a maneira como o contexto global e a estratégia do sistema de avaliação afetam a aceitação e o uso dos resultados por parte dos estabelecimentos educacionais. Além disto, seria relevante comparar o uso das informações entre países nos quais a divulgação das mesmas se dá em

<sup>9</sup> CARNOY, M., 2000, p. 18, em inglês no original.

um contexto de capacitação – o que, em princípio, parece mais eficaz, ainda que também mais oneroso – e países em que o sistema de avaliação se limita a distribuir os relatórios entre as escolas.

Em todo caso, o uso dos resultados dos sistemas de avaliação desenvolvidos até o momento com fins de “prestação de contas” requer uma cuidadosa análise dos riscos que implica, como adverte Linn:

*“Não deveríamos esperar que provas de custo baixo desenhadas para outros propósitos de ‘baixo risco’ dêem conta das pressões que os sistemas de prestação de contas de ‘consequências fortes’ lhes impõem atualmente (...) Sistemas de avaliação que são monitores extremamente úteis perdem boa parte de sua credibilidade para tal propósito quando incorporam ‘consequências fortes’. Os efeitos negativos não desejados do uso das avaliações para fins de prestações de contas de ‘consequências fortes’ freqüentemente superam os efeitos positivos desejados.”<sup>10</sup>*

## **Capítulo II.**

### **Percentuais, Médias e Pontuações: as Métricas Empregadas Para Informar os Resultados**

Como uma segunda abordagem aos relatórios das avaliações nacionais, no presente capítulo serão analisadas as “métricas” que eles utilizam, isto é, os tipos de dados que são apresentados para descrever os resultados em nível nacional.

Sete “métricas” têm sido empregadas nos diversos relatórios nacionais para caracterizar a qualidade dos sistemas educativos e informar os resultados das avaliações:

- percentual médio de respostas corretas;
- média das pontuações “brutas”;
- pontuações padronizadas;

- pontuações de “Teoria de Resposta ao Item”;
- percentual de alunos que atingem certo nível de pontuação ou de desempenho;
- percentual de alunos que demonstram domínio de objetivos específicos;
- uma transformação de algumas das métricas anteriores em uma qualificação numérica expressa na mesma escala que habitualmente se emprega no sistema para qualificar os alunos.

É importante compreender o que significam as diferentes métricas ou tipos de dados que são apresentados nos relatórios nacionais por duas razões principais:

- porque está em jogo a “compreensibilidade” das informações para os destinatários;
- porque está em jogo a “utilidade” das informações apresentadas, em função das necessidades dos destinatários. Por exemplo, caso se espere que as informações sirvam aos professores para revisar o modo como encaram o ensino, é necessário definir se basta informar o percentual de respostas corretas nas provas em nível nacional, ou se é preciso informá-lo por região.

Em outro plano, é necessário indagar se o fato de informar-se ao público que uma média de 58% das respostas estavam corretas atenderia àquela que parece ser a demanda principal do público: saber se os alunos estão aprendendo o que se esperava deles ao final de determinados níveis de escolarização.<sup>11</sup>

Ao longo deste capítulo explica-se de forma sintética o significado de cada uma das métricas empregadas nos distintos países da região, incluindo-se um modelo de tabela e uma rápida revisão de suas potencialidades e seus problemas. O leitor interessado em mais

<sup>10</sup> LINN 2000, pp.12-14, em inglês no original.

<sup>11</sup> Este tema envolve saber se os dados foram construídos dentro de um enfoque referido a normas ou dentro de um enfoque baseado em critérios, questão tratada em mais detalhe em RAVELA, P. (editor) e outros, 2000 e que será analisada ao final deste capítulo.

detalhes pode consultar a versão completa deste texto na página do PREAL na Web.

## **II. 1. Percentual médio de respostas corretas e/ou médias de pontuações brutas**

Calcula-se o percentual médio de respostas corretas calculando primeiro que percentual dos alunos respondeu corretamente a cada uma das questões da prova, e depois a média desses percentuais para todas as questões da prova. Em termos matemáticos, esta “métrica” é equivalente à média das pontuações brutas<sup>12</sup> de todos os estudantes.

---

<sup>12</sup> Em uma prova de múltipla escolha, as pontuações “brutas” são as que se calculam para cada estudante a partir do número de questões que respondeu corretamente. Se um prova contiver 40 questões, a pontuação “bruta” variará entre 0 e 40. Se um aluno tiver respondido corretamente a 22 questões, sua pontuação bruta será 22.

Argentina, Chile e Paraguai são os países que empregam o percentual médio de respostas corretas em cada prova como métrica principal de seus relatórios. Na figura 3 incluí-se, a título ilustrativo, uma tabela retirada de um dos relatórios argentinos.

Através desta métrica, o relatório permite uma visão muito geral sobre os resultados alcançados. Informa sobre o domínio de certas matérias no sistema educativo, mas não sobre o que os alunos, considerados individualmente, são capazes de fazer, pois as unidades de análise são as questões, não os alunos.

Um resultado de 55% de respostas corretas pode resultar, tanto do fato de que a maioria dos alunos localizou-se em torno de tal média, como de uma forte dispersão em situações extremas: alunos com alto domínio dos conteúdos avaliados, de um lado, e alunos cujas pontuações foram muito baixas, do outro.

Como a escolha das questões que integrarão a prova em geral privilegia aquelas que tenham nível de dificuldade médio, e não partem de uma verificação substantiva de qual nível de aprendizado se espera que os alunos tenham atingido, o resultado na realidade não diz nada sobre o grau em que os alunos são capazes de fazer o que deles se espera.

Em outras palavras, os resultados usualmente observados, entre 50% e 60% de respostas corretas, na realidade são determinados *a priori*, quando se escolhem as questões, mas não significam que se esteja atingindo este percentual do que se espera que os alunos tenham aprendido.

Este tipo de métrica pode ser empregada de forma válida para comparações entre distintos setores do sistema (departamentos, regiões, escolas, etc.), analisando-se quais os que estão acima ou abaixo da média nacional, bem como, a distância entre eles, como é feito na Argentina. (figura 3).

Finalmente, a informação através de percentuais médios de respostas corretas também pode ser empregado para informar sobre mudanças no sistema ao longo dos anos, como foi feito no Chile e na Argentina(\*).

(\*) Para que esta informação permita apreciar de forma válida a evolução dos resultados do sistema educativo ao longo do tempo, é necessário que as provas estejam adequadamente equiparadas em nível de dificuldade com as da primeira medição, que meçam os mesmos conteúdos e/ou competências e que tenham a mesma estrutura em termos de quantidade de questões e dos pesos das diferentes competências na pontuação total. Nos relatórios públicos analisados destes países não há informações sobre os procedimentos adotados para assegurar a equivalência dos instrumentos de medição utilizados nos diversos anos.

**Figura 3**  
**ARGENTINA - VALORES MÉDIOS DE RENDIMENTO EM LÍNGUA POR JURISDIÇÃO E NÍVEL**

| <b>Nivel Primario - 6° Grado Lengua</b> |                        |   |
|---|------------------------|---|
| <b>Muestra 1996</b>                     |                        |   |
| <b>Jurisdicción</b>                     | <b>Lengua 6° Grado</b> |   |
|   | <b>Muestra 1996</b>    | <b>Relación con respecto al valor medio</b> |
| Capital Federal                         | 72.39%                 | 1.24  |
| Mendoza                                 | 67.63%                 | 1.16  |
| Río Negro                               | 66.28%                 | 1.14  |
| Neuquén                                 | 65.37%                 | 1.12  |
| La Pampa                                | 60.58%                 | 1.04  |
| Chaco                                   | 60.37%                 | 1.04  |
| Córdoba                                 | 58.77%                 | 1.01  |
| Santa Cruz                              | 58.69%                 | 1.01  |
| <b>MEDIA NACIONAL</b>                   | <b>58.19%</b>          | <b>1.00</b>                                 |
| Buenos Aires                            | 57.88%                 | 0.99  |
| Santa Fe                                | 57.87%                 | 0.99  |
| Entre Ríos                              | 57.26%                 | 0.98  |
| Gran Buenos Aires                       | 56.98%                 | 0.98  |
| Tierra del Fuego                        | 56.87%                 | 0.98  |
| Misiones                                | 55.83%                 | 0.96  |
| Chubut                                  | 55.41%                 | 0.95  |
| Salta                                   | 54.71%                 | 0.94  |
| Corrientes                              | 54.12%                 | 0.93  |
| San Juan                                | 53.79%                 | 0.92  |
| Jujuy                                   | 52.16%                 | 0.90  |
| La Rioja                                | 50.14%                 | 0.86  |
| Tucumán                                 | 49.04%                 | 0.84  |
| Formosa                                 | 48.94%                 | 0.84  |
| Santiago del Estero                     | 47.74%                 | 0.82  |
| Catamarca                               | 46.48%                 | 0.80  |

Fonte: Ministerio de Cultura y Educación de la Nación, Argentina, 1997. Operativo Nacional de Evaluación de Aprendizajes 1996; página 33.

## **II. 2. Percentuais de alunos que dominam certos objetivos ou que atingem níveis pré-determinados de desempenho**

Ao informar os resultados, Costa Rica, Equador e Uruguai adotam um enfoque substancialmente diferente daquele dos três países anteriores, utilizando como métrica principal o percentual de indivíduos (alunos) que atingem níveis de domínio pré-definidos nas competências, destrezas e/ou conteúdos avaliados. As principais diferenças em relação ao enfoque descrito anteriormente são:

- os resultados não são unicamente descritivos, implicando uma avaliação quanto a se os alunos alcançaram ou não o que se esperava deles;
- os resultados, por conseguinte, podem ser interpretados, não apenas em termos de comparações entre indivíduos e grupos, como também em termos do grau em que as metas de ensino estão sendo alcançadas.

Costa Rica e Equador elaboram suas provas dentro do enfoque de avaliações referidas a critérios. Para tanto, em cada disciplina avaliada define-se um conjunto de dez a quatorze objetivos ou destrezas específicos que sejam considerados prioritários. O “domínio” de cada objetivo ou destreza por parte dos alunos é avaliado através de várias questões referidas ao mesmo: três, no caso da Costa Rica e quatro no do Equador. No primeiro país considera-se que um aluno domina determinado objetivo se responde corretamente a pelo menos a duas das três questões referidas ao mesmo. No caso do equador, quando responde corretamente ao menos a três das quatro questões.

Os relatórios de resultados desses países empregam como “métrica” principal o percentual de alunos que domina cada objetivo. Nos relatórios, os resultados aparecem desagregados por destreza ou objetivo, dando assim uma visão detalhada do que os alunos são capazes de fazer. Na figura 4 apresenta-se, a título de ilustração, uma tabela de um relatório nacional da Costa Rica.

**Figura 4**  
**COSTA RICA. Percentual de estudantes que domina cada objetivo por região. Sexta série.**  
**Espanhol. IIMEC-MEP. Outubro-novembro de 1996**

| OBJETIVO | AMOSTRAL | CENTRAL | BRUNCA | CHOROTEGA | HUETAR NORTE | PACÍFICO CENTRAL | HUETAR ATLÁNTICA |
|----------|----------|---------|--------|-----------|--------------|------------------|------------------|
| 1        | 69,9     | 72,0    | 67,5   | 60,6      | 75,0         | 65,1             | 64,2             |
| 2        | 69,9     | 87,8    | 87,6   | 78,8      | 85,3         | 80,2             | 85,5             |
| 3        | 93,7     | 95,5    | 91,6   | 85,0      | 92,6         | 92,5             | 90,5             |
| 4        | 81,4     | 83,4    | 76,7   | 76,3      | 82,4         | 81,6             | 72,6             |
| 5        | 67,5     | 69,6    | 71,3   | 51,3      | 70,6         | 64,9             | 62,6             |
| 6        | 78,7     | 80,9    | 76,9   | 66,3      | 75,0         | 77,9             | 77,1             |
| 7        | 93,6     | 95,2    | 91,0   | 86,3      | 88,2         | 93,6             | 92,7             |
| 8        | 91,7     | 93,8    | 88,6   | 85,6      | 90,4         | 89,5             | 85,5             |
| 9        | 93,9     | 65,1    | 92,9   | 85,6      | 91,1         | 91,3             | 97,1             |
| 10       | 61,0     | 63,8    | 58,7   | 48,1      | 55,1         | 56,4             | 58,1             |
| 11       | 78,4     | 84,1    | 66,9   | 63,1      | 72,8         | 75,0             | 65,4             |
| 12       | 81,4     | 84,2    | 77,8   | 62,5      | 80,1         | 81,4             | 75,4             |
| 13       | 94,5     | 95,9    | 91,6   | 88,8      | 94,1         | 90,7             | 95,0             |
| 14       | 80,1     | 83,7    | 80,8   | 64,4      | 74,3         | 77,3             | 71,5             |

**Fonte:** Ministerio de Educación Pública – Universidad de Costa Rica/Instituto de Investigaciones para el Mejoramiento de la Educación Costarricense (IIMEC), Costa Rica, 1997. *Informe Nacional sobre el Desarrollo, Validación y Aplicación de las pruebas de Diagnóstico de conocimientos. Octubre-noviembre de 1996;* pág. 39.

Este modo de informar apresenta uma dificuldade na hora de apresentar uma visão global do desempenho dos alunos no conjunto da prova, ou dos objetivos e destrezas de determinada disciplina. Para resolvê-la, tanto a Costa Rica como o Equador recorrem, por caminhos diversos, a uma média de pontuações brutas, semelhante à dos países mencionados anteriormente. No caso da Costa Rica, informa-se uma pontuação média do conjunto da prova, calculada como percentual da pontuação total. O Equador utiliza uma qualificação de 20 pontos – que é a escala que rege o sistema para avaliação do aluno que se constrói, como nos casos anteriores, com base no percentual médio da pontuação total da prova.

A modalidade adotada no Uruguai para comunicar os resultados é similar, quanto à métrica: o dado principal é o percentual de alunos que alcança determinado padrão de

rendimento ou categoria de aprendizado definido a priori e não é meramente descritivo, implicando uma avaliação do aprendizado dos alunos. No Uruguai, esta categoria de aprendizado é denominada “suficiência” e é estabelecida para o conjunto da prova – neste sentido é diferente do que fazem Equador e Costa Rica – e para cada uma das grandes “competências” em que cada prova está estruturada.

Para estabelecer a “suficiência”, definem-se “pontos de corte” na pontuação bruta das provas, pontos estes que não estão referidos unicamente à distribuição normal da população, mas a um parecer técnico de especialistas e professores sobre o que os alunos deveriam ser capazes de realizar ao concluir o curso primário. As questões com alto ou baixo grau de dificuldade não são descartadas. São mantidas na medida em que aferem aprendizados definidos como fundamentais para um aluno que termina o

primário. À luz dos resultados da prova piloto e a partir do parecer de árbitros especialistas, selecionam-se as questões de modo tal que, como mínimo, se espere que todos os alunos respondam a 14 de um total de 24 questões (o que equivale a

aproximadamente 60% da pontuação bruta total). Este é o nível de “suficiência” para o conjunto da prova. Os relatórios concentram-se em informar o percentual de alunos que alcançou este nível de desempenho (ver figura 5).

**Figura 5**  
**URUGUAI. Resultados em Língua e Matemática, Sexta Série,**  
**Por faixa de pontuação em 1996 e 1999**

| Percentual de alunos:                                | LÍNGUA      |             | MATEMÁTICA  |             |
|--|-------------|-------------|-------------|-------------|
|  | 1996        | 1999        | 1996        | 1999        |
| Com desempenho altamente satisfatório (20-24 pontos) | 15,8        | 14,0        | 6,8         | 9,0         |
| Com desempenho satisfatório (14 -19 pontos)          | 41,3        | 47,3        | 27,8        | 31,8        |
| <b>SUFICIENTES</b>                                   | <b>57,1</b> | <b>61,3</b> | <b>34,6</b> | <b>40,8</b> |
| Com desempenho insatisfatório (7 - 13 pontos)        | 37,7        | 35,3        | 54,5        | 50,9        |
| Com desempenho muito insatisfatório (0 - 6 pontos)   | 5,2         | 3,4         | 10,9        | 8,3         |
| <b>INSUFICIENTES</b>                                 | <b>42,9</b> | <b>38,7</b> | <b>65,4</b> | <b>59,2</b> |
| Totais   | 100,0       | 100,0       | 100,0       | 100,0       |

**Fonte:** Administración Nacional de Educación Pública - Uruguay / Unidad de Medición de Resultados Educativos. *Evaluación Nacional de Aprendizajes en Lengua y Matemática. 6<sup>to</sup>. año Enseñanza Primaria - 1999. Primer Informe*; pág. 25.

Assim como nos casos do Chile e da Argentina, o Uruguai apresentou seus resultados de 1999 em tabelas comparativas com as medições anteriores<sup>13</sup>.

Um dos pontos fracos dos relatórios uruguaios é que não estão explicitamente fundamentadas as razões pelas quais os pontos de corte são estes, não outros, e o que é que se supõe que marque a diferença, em

termos de conhecimentos e competências entre um aluno “suficiente” e um que não o é.<sup>14</sup>

### II .3. Pontuações padronizadas

Na apresentação de seus resultados a Bolívia utiliza pontuações “padronizadas”, que em estatística são denominadas pontuações “T”. Trata-se de uma transformação matemática da pontuação bruta

<sup>13</sup> As provas aplicadas em 1999 são formas equivalentes às utilizadas em 1996 e o procedimento de elaboração das mesmas está descrito no relatório público **ANEP/UMRE, 1999; Evaluación Nacional de Aprendizajes en Lengua y Matemática. 6<sup>to</sup>. año Enseñanza Primaria-1999. Primer Informe**; pp. 3 a 9.

<sup>14</sup> O estabelecimento de "pontos de corte" para definir categorias de desempenho nunca é uma atividade "objetiva"; é sumamente complexa e fortemente “vinculada” aos pontos de vista dos especialistas que determinam tais pontos de corte.

de cada aluno a partir da distribuição das pontuações brutas verificada na população. A pontuação atribuída a cada indivíduo representa, assim, sua distância em relação à média da população, em unidades de desvio padrão. Para tanto, emprega-se – na Bolívia – uma média de 50 pontos e um desvio padrão de 10. Isto significa que as pontuações 40 e 60 correspondem aos alunos que se localizam na faixa do desvio padrão, abaixo ou acima da média nacional. Os alunos abaixo de 40 tiveram resultados significativamente inferiores aos da maioria, enquanto os que estão acima de 60 são os que tiveram resultados superiores à maioria do universo avaliado.

Este modo de informar é similar ao dos percentuais médios de respostas corretas empregado na Argentina, no Chile e no Paraguai, destacando-se principalmente por permitir a possibilidade de comparação de resultados entre indivíduos ou grupos. A transformação em pontuações “T” permite afinar estatisticamente essas comparações, mas é menos compreensível para o público não especializado.

No relatório boliviano são estabelecidos pontos de corte nos 45 e 55 pontos da escala padronizada. A partir deles definem-se três grandes categorias de desempenho dos alunos:

- *satisfatório*, quando o aluno tem mais de 55 pontos
- *regular*, quando a pontuação do aluno se situa entre 45 e 55
- *risco*, quando o aluno não alcança os 45 pontos

Os relatórios mostram logo o percentual de alunos em cada uma destas três categorias. Sem embargo, é preciso ter em mente que esta classificação não é absoluta, mas relativa: *risco* significa que um aluno está abaixo da média nacional e na parte inferior da distribuição, mas nada revela a respeito de seu nível de aprendizado dos objetivos curriculares. Do mesmo modo, *satisfatório* indica que o aluno está acima do rendimento da maioria dos alunos, mas não implica

necessariamente que tenha atingido de forma satisfatória os objetivos curriculares. Vale dizer, trata-se de um padrão referido à distribuição populacional, o que é substancialmente diferente do enfoque adotado na Costa Rica, no Equador e no Uruguai.

#### II. 4. Teoria de Resposta ao Item

O Brasil foi o primeiro país da região a utilizar a Teoria de Resposta ao Item (TRI) para construção das pontuações de suas avaliações nacionais<sup>15</sup>. A TRI é um modelo matemático que permite estimar a capacidade dos indivíduos em uma determinada área ou disciplina a partir da premissa de que ela é unidimensional. Vale dizer: presume-se, por exemplo, que os alunos tenham uma capacidade ou competência para a matemática que determina a probabilidade de que determinado aluno realize adequadamente as diferentes atividades incluídas no banco de questões.

A explicação detalhada da TRI excede em muito os limites do presente trabalho, mas basta dizer que ela tem algumas vantagens sobre o enfoque clássico:

- permite pôr questões e alunos em uma mesma escala;
- permite estimativas mais precisas das mudanças ao longo do tempo, mediante equiparação das pontuações;
- permite estimar uma medida da capacidade dos alunos que leva em conta a dificuldade das questões, isto é, as questões mais difíceis têm peso maior na determinação da pontuação individual.

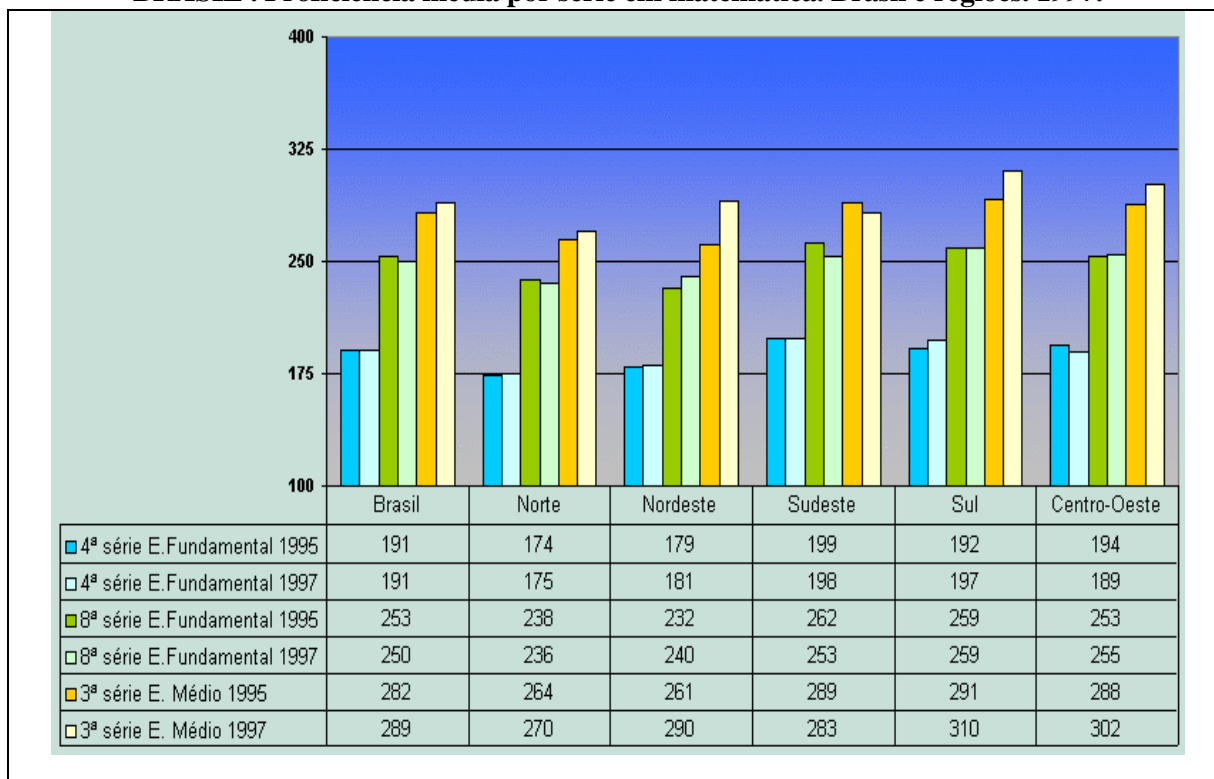
As pontuações de TRI não têm uma métrica especificada. Normalmente toma-se arbitrariamente uma média de 250 pontos e um desvio padrão de 50, ou uma média de 500 pontos e um desvio padrão de 100. No

<sup>15</sup> A partir de 1999 Chile e Argentina começaram a informar seus resultados com este modelo que, por outro lado, é o mesmo adotado nos estudos internacionais. O Peru, que publicou seus primeiros resultados em 2000 também o utiliza.

caso do SAEB (Sistema de Avaliação da Educação Básica) brasileiro, adota-se uma média de 250 pontos, que corresponde à média nacional de 1997 dos alunos da 8ª série. Como se pode ver na figura 6, a partir das premissas da TRI é possível construir uma escala única de pontuações, desenvolvendo matrizes de referência para populações de distintos níveis de escolaridade: no caso do Brasil, 4ª e 8ª séries

do ensino fundamental (educação básica) e 3ª do ensino médio. Isto permite comparar as médias de “proficiência” em cada disciplina entre os diversos níveis do sistema educativo e entre as regiões do país, e entre os vários anos, colocando todos os níveis em uma mesma tabela.

**Figura 6**  
**BRASIL . Proficiência média por série em matemática. Brasil e regiões. 1997.**



**Fonte:** Ministério da Educação - Governo Federal. Instituto Nacional de Estudos e Pesquisas Educacionais (INEP); <http://www.inep.gov.br/saeb/saeb97/saeb97.htm>. SAEB/97.Estudo Comparativo.

Do mesmo modo que nas pontuações padronizadas adotadas na Bolívia, um dos problemas do uso da TRI é a dificuldade em comunicar ao público e aos professores como se chega a determinar as pontuações e o que elas significam. Dificilmente terão sentido para o leitor não especializado, enquanto não existir uma tradição ou cultura em torno de sua interpretação.

Um modo de enfrentar esta dificuldade, cada vez mais adotado nos estudos internacionais e nos relatórios da *National Assessment of Educational Progress NAEP* (Avaliação Nacional do Progresso Educacional) do *National Center for Education Statistics* (Centro Nacional de Estatísticas da Educação) dos Estados Unidos, é fazer uma clara descrição do que são capazes de fazer os alunos que se encontram em determinados pontos da escala

de pontuações. No caso do SAEB, foram estabelecidos pontos de corte para cada 75 pontos, explica-se de que são capazes os alunos em cada um desses “níveis de proficiência” (figura 7), e informa-se o percentual de alunos que alcançou ou superou cada nível. A descrição de tais níveis é feita a partir das questões que conseguem responder os indivíduos cuja pontuação se localiza em torno de um dos pontos de corte, e do que **não** conseguem responder os indivíduos cuja pontuação fica em torno do ponto de corte anterior.

Nos estudos internacionais, este modo de informar é empregado de maneira estritamente descritiva, isto é, descreve-se o que os alunos são capazes de fazer em cada faixa da escala, mas não se responde à pergunta: isto é o que deveriam saber? É mais, ou menos? Estamos bem, regulares, ou mal? Obviamente, ater-se a um plano descritivo é adequado para os estudos internacionais, dado que seria quase impossível chegar a um consenso em torno de critérios comuns a todos os países – tarefa que é indispensável no nível nacional.

**Figura 6**  
**BRASIL/SAEB 1997 - ESCALA DE PROFICIÊNCIA EM MATEMÁTICA**

| <b>Exemplos de Desempenho</b>   | <b>NÍVEL</b> | <b>Resultados do SAEB/97</b>   |
|---|--------------|--|
| <ul style="list-style-type: none"> <li>▪ Os alunos reconhecem o valor de cédulas e moedas.</li> <li>▪ Lêem a hora em relógios digitais e analógicos e sabem que uma hora tem 60 minutos.</li> <li>▪ Resolvem problemas simples de adição e subtração com números naturais.</li> </ul>   | <b>175</b>   | <p><b>No nível 175 ou acima dele, encontram-se:</b></p> <ul style="list-style-type: none"> <li>▪ 56% dos alunos da 4ª série do Ensino Fundamental;</li> <li>▪ 95% dos alunos da 8ª série do Ensino Fundamental;</li> <li>▪ 100% dos alunos da 3ª série do Ensino Médio.</li> </ul> |
| <ul style="list-style-type: none"> <li>▪ Os alunos reconhecem polígonos e quadriláteros.</li> <li>▪ Estabelecem relações entre os vários valores de cédulas e moedas e resolvem situações ao pagar e receber troco, ainda que não saibam operar com decimais.</li> <li>▪ São capazes de multiplicar e dividir, bem como identificar unidades, dezenas e centenas.</li> <li>▪ Resolvem problemas que envolvem mais de uma operação.</li> <li>▪ Adicionam e subtraem frações de um mesmo denominador e conhecem números naturais na forma fracionária.</li> <li>▪ Interpretam gráficos de barras e de sector e identificam o gráfico mais adequado pra representar uma determinada situação.</li> </ul> | <b>250</b>   | <p><b>No nível 250 ou acima dele, encontram-se:</b></p> <ul style="list-style-type: none"> <li>▪ 11% dos alunos da 4ª série do Ensino Fundamental;</li> <li>▪ 48% dos alunos da 8ª série do Ensino Fundamental;</li> <li>▪ 87% dos alunos da 3ª série do Ensino Médio.</li> </ul>  |
| <ul style="list-style-type: none"> <li>▪ Os alunos classificam sólidos geométricos em corpos redondos e poliedros.</li> <li>▪ Interpretam resultados de medidas de longitude, massa, tempo e capacidade.</li> <li>▪ Identificam, comparam e ordenam números racionais (nas formas fracionária e decimal) bem como números inteiros.</li> <li>▪ Interpretam linguagem algébrica e resolvem equações e sistemas de equações de primeiro grau.</li> </ul>  | <b>325</b>   | <p><b>No nível 325 ou acima dele, encontram-se:</b></p> <ul style="list-style-type: none"> <li>▪ 8% dos alunos da 8ª série do Ensino Fundamental;</li> <li>▪ 32% dos alunos da 3ª série do Ensino Médio.</li> </ul>  |
| <ul style="list-style-type: none"> <li>▪ Os alunos resolvem problemas que envolvem ponto, reta, circunferência e suas relações.</li> <li>▪ Estabelecem relações e fazem conversões entre frações ordinárias e números decimais.</li> <li>▪ Resolvem problemas que envolvam equações e inequações simples de primeiro e segundo graus e sistemas de primeiro grau.</li> <li>▪ Conhecem os princípios básicos de polinômios e realizam operações elementares entre eles.</li> <li>▪ Conhecem as propriedades básicas de expoentes e logaritmos.</li> </ul>  | <b>400</b>   | <p><b>No nível 400 ou acima dele, encontram-se:</b></p> <ul style="list-style-type: none"> <li>▪ 5% dos alunos da 3ª série do Ensino Médio.</li> </ul>   |

Observação: Os níveis 325 e 400 demonstram o domínio de habilidades e conteúdos mais complexos que não correspondem à 4ª série do Ensino Fundamental, Por isto não se apresentam resultados para os alunos desta série.

**Fonte:** Ministério da Educação - Governo Federal. Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), Brasil, 1998; *Como está a Educação Básica no Brasil?*

O SAEB enfrentou este desafio da seguinte maneira: a partir dos resultados descritivos do que são capazes de fazer os

alunos que se encontram em diferentes pontos da escala, grupos de especialistas definem em que momento da escolaridade os alunos

deveriam ter atingido cada um destes níveis, com o que se estabelece um critério de

valoração que permite responder às perguntas especificadas no parágrafo anterior (figura 8).

**Figura 8**  
**BRASIL. Relação entre níveis de proficiência e ciclos dos níveis de ensino**

| Nível de proficiência - escala SAEB/97 | Matemática                                     | Língua Portuguesa                              | Ciências (Física, Química e Biologia)          |
|--|--|--|--|
|  | <b>Ciclo e nível de ensino</b>                 | <b>Ciclo e nível de ensino</b>                 | <b>Ciclo e nível de ensino</b>                 |
| 100                                    | Não significativo                              | Até a metade do 1º ciclo do Ensino Fundamental | Até a metade do 1º ciclo do Ensino Fundamental |
| 175                                    | Até a metade do 1º ciclo do Ensino Fundamental | Até o fim do 1º ciclo do Ensino Fundamental    | Até o final do 1º ciclo do Ensino Fundamental  |
| 250                                    | Até o final do 1º ciclo do Ensino Fundamental  | Até o final do 1º ciclo do Ensino Fundamental  | Até a metade do 2º ciclo do Ensino Fundamental |
| 325                                    | Até o final do 2º ciclo do Ensino Fundamental  | Até o final do Ensino Médio                    | Até o final do 2º ciclo do Ensino Fundamental  |
| 400                                    | Até o final do Ensino Médio                    | Após o final do Ensino Médio                   | Até o final do Ensino Médio                    |

**Fonte:** Ministério da Educação - Governo Federal. Instituto Nacional de Estudos e Pesquisas Educacionais (INEP). *SAEB 97 - Primeiros Resultados*.

Em outras palavras, o leitor pode comparar o nível de escolaridade em que o aluno supostamente deveria ter alcançado um certo nível de pontuação ou de “proficiência”, com a proporção dos alunos de tal nível de ensino que efetivamente o conseguiu<sup>16</sup>.

Este é um modo eficaz de apresentar as informações, dado que permite ao leitor atribuir algum sentido à escala de pontos em termos do que são capazes de fazer os estudantes que alcançaram cada um dos níveis de referida escala e fazer uma avaliação da situação.

<sup>16</sup> A este respeito, é preciso observar – e o relatório do SAEB o indica explicitamente – que no caso do Brasil esta tarefa é extremamente complexa em virtude da diversidade curricular existente no país. Idealmente este tipo de níveis de “proficiência” ou de desempenho deveriam poder ser estabelecidos a partir de um currículo específico, e não como mínimo denominador comum de vários currículos distintos.

## II. 5. Qualificações

Muitos países transformam os dados recolhidos nas avaliações em uma escala de qualificações similar àquela geralmente utilizada para qualificar os alunos no sistema educativo. Esta prática se justifica com o argumento de que pais e educadores estão acostumados a esta qualificação:

*“Esta qualificação tem um significado genérico e, ao mesmo tempo, limitado, já que não informa que destreza foi mais ou menos dominada pelo aluno de determinada série, que é o assunto que efetivamente importa nas medições do APRENDO. No entanto, este relatório apresenta as referidas qualificações porque são de uso comum na educação nacional”* (Ministerio de Educación y Cultura del Ecuador – EB/PRODEC, 1998)

Isto, no entanto, pode gerar algumas confusões que merecem ser mencionadas. A primeira, e mais óbvia, é justamente a de assemelhar os resultados das avaliações às qualificações de uso habitual no sistema, quando seus sentido e significado são completamente diferentes. Assim, melhor

seria trabalhar no sentido de ajudar a compreender que as avaliações são diferentes, que não têm fins de certificação e que não podem ser comparadas às que os professores realizam em aula.

Por outro lado, na maioria dos países que não estabelecem critérios ou padrões de aprendizado, o uso da qualificação favorece a imagem de que se trata de um resultado avaliativo do que o sistema educativo ou uma determinada região está conseguindo (por definição, as qualificações têm um ponto de corte que implica a divisão entre aprovado e reprovado). Esta interpretação é totalmente inadequada quando a qualificação se estabelece como uma transformação matemática da média de pontuações brutas ou do percentual de respostas corretas a partir de provas em que as questões tenham sido selecionadas propositalmente para que sua dificuldade média se localize entre .30 e .80.

Como já foi indicado, quando as provas são desenhadas desta

maneira seus resultados só devem ser interpretados em termos comparativos, evitando-se qualquer interpretação em termos de alunos reprovados ou de aprovação/reprovação do trabalho do sistema educativo, ou de determinada gestão.

## II. 6. Resultados descritivos ou padrões de aprendizado: o que é que o público e os professores desejam saber?

Na Figura 9 apresenta-se um resumo das "métricas" que são empregadas nos relatórios dos diferentes países da região.

**Figura 9**

### Métricas empregadas pelos distintos países para comunicação dos resultados

| Países  | Perc. médio de respostas corretas | Média de pontuações brutas | Pontuações padronizadas | Pontuações "teoria de resposta à questão" | Perc. de alunos por níveis de desempenho | Per. de alunos que domina os objetivos | Qualificações |
|---------|-----------------------------------|----------------------------|-------------------------|---|--|--|---------------|
| ARGENTI |                                   |                            |                         |   |  |  |               |
| BOLÍVIA |                                   |                            | <b>Pontuação</b>        |   |  |  |               |
| BRASIL  |                                   |                            |                         |   |  |  |               |
| COSTA   |                                   |                            |                         |   |  |  |               |
| CHILE   | <b>Até 1997</b>                   |                            |                         | <b>Desde 1998</b>                         |  |  |               |
| EQUADO  |                                   |                            |                         |   |  |  |               |
| PARAGU  |                                   |                            |                         |   |  |  |               |
| PERU    |                                   |                            |                         |   |  |  |               |
| URUGUA  |                                   |                            |                         |   |  |  |               |

Nota: As células mais escuras indicam qual a métrica central utilizada nos relatórios de cada país.

Ao longo deste capítulo pudemos apreciar que, por trás das métricas ou "tipos de dados" usados para apresentar os resultados das avaliações nacionais de aprendizado, há diversos significados e implicações sobre os quais é preciso informar a opinião pública e a dos professores de modo que possam interpretar e usar adequadamente os dados.

A principal alternativa implicada é a seguinte:

- ou limitar-se a descrever o que os alunos conseguem fazer em função das tarefas propostas e a estabelecer comparações entre diferentes setores do sistema educativo, ou entre anos, para chegar a conclusões do tipo "os aprendizados em Língua são melhores nas escolas urbanas que nas rurais",

“em tal estado em comparação com tal outro”, ou “os resultados em Matemática melhoraram – ou pioraram – em comparação com os de dois anos atrás”;

- ou estabelecer padrões ou critérios de referência para os desempenhos esperados e exigidos de todos os alunos que terminam determinada série ou nível, o que permite, além dos tipos de conclusões indicadas em “a”, outras do tipo “tal proporção dos estudantes não atingiu os níveis esperados de desempenho em leitura na 3ª série”, ou “tal proporção dos estudantes domina o objetivo ou conteúdo “x” de Matemática da 6ª série”.

Não resta dúvida de que as conclusões do primeiro tipo são úteis, fundamentalmente, para que se tenha uma visão das desigualdades verificadas no interior do sistema educativo e da evolução dos resultados ao longo do tempo. No entanto, as expectativas da opinião pública e das famílias, e o modo como tendem a interpretar os resultados, parecem estar exigindo conclusões do segundo tipo. Além disto, conclusões deste último teor provavelmente terão efeito mais forte sobre os professores, em termos de sinalizar para eles o que se espera que todos os alunos saibam e sejam capazes de fazer ao final de cada série ou ciclo de ensino.

Este último exige um considerável esforço para traduzir o currículo – que normalmente, em nossos países, é formulado em termos bastante genéricos – em especificações mais claras e precisas do tipo de desempenhos que se espera dos alunos e incorporar metodologias específicas para estabelecimento de critérios sobre os níveis de aprendizado exigidos ao final de determinadas séries ou ciclos. Em países onde não haja um currículo nacional comum, a tarefa, obviamente, implicará um esforço maior de construção de consensos sobre o que pode ser exigido de todos.

### Capítulo III.

#### Comparando o Incomparável: a Parábola das Estradas

##### III. 1 Modos típicos de desagregar as informações nos relatórios nacionais

Uma das principais deficiências dos relatórios de resultados das avaliações nacionais nos países da região é a falta de contextualização sociocultural de tais resultados. Em praticamente todos os países são adotados os seguintes níveis de desagregação:

- resultados segundo as divisões políticas e/ou geográficas do país (por estado, província, departamento ou região);
- resultados de escolas urbanas e rurais;
- resultados segundo a forma de administração (escolas públicas ou fiscais, escolas subvencionadas, escolas particulares ou privadas).

Estes níveis de desagregação, que atendem às diferenciações mais “visíveis” dos sistemas educativos em subsistemas, são relevantes na medida em que permitem ver as diferenças nos aprendizados dos alunos entre tais subsistemas e trazem a público certas desigualdades na distribuição dos conhecimentos. Apesar disto, o fato de serem, praticamente, as únicas formas de desagregação dos dados implica uma simplificação excessiva da realidade, na medida em que ocorrem dois fenômenos:

- alguns destes agregados incorporam enormes heterogeneidades. Por exemplo, o agregado “escolas urbanas” inclui toda a gama de situações socioculturais, desde escolas de setores de classe média e alta, até escolas da periferia. O mesmo ocorre com a maioria das “jurisdições” políticas ou regiões geográficas em que se divide um país;
- em outros casos, os diferentes tipos de escola têm alunos que pertencem majoritariamente a determinados setores sociais, e não a uma população representativa do conjunto: tipicamente, as escolas rurais – que trabalham com crianças originárias de

famílias com trajetória mínima no sistema educativo – e as escolas privadas – que recrutam a maioria de seus alunos entre as classes média e alta da sociedade.

O problema se apresenta quando, com base nestas formas de desagregação dos resultados, derivam-se julgamentos ou conclusões sobre diferenças em eficácia de uns e outros tipos de escolas ou jurisdições, sem levar em conta que trabalham com alunos diferentes.

Em geral, como o demonstra a abundante pesquisa empírica<sup>17</sup>, as diferenças de resultados obedecem principalmente às diferenças socioculturais entre as populações com as quais trabalham os distintos setores ou instituições do sistema educativo, antes que a diferenças em sua eficácia pedagógica.

O que foi dito acima não implica afirmar que todas as diferenças de resultados possam ser explicadas a partir das diferenças no contexto social, mas que, para se fazerem afirmações válidas a respeito da “eficácia” pedagógica, é imprescindível controlar o efeito do contexto sociocultural.

### III. 2. A parábola das estradas

Para explicar melhor este problema é útil recorrer à seguinte metáfora: na República Argentina há uma estrada denominada “rota do deserto”. É a Estrada Nacional No. 20, que corta a província de La Pampa na direção leste-oeste. Ela atravessa uma paisagem

semidesértica, pedregosa, com arbustos e totalmente plana. A estrada tem mais de 200 km em linha reta, sem nenhuma curva ou lombada. O maior perigo é a monotonia da paisagem, que leva o motorista a dormir na direção. Por este motivo, ao longo de toda a estrada há sinais advertindo para a necessidade de descansar, e a cada 60 quilômetros foram localizados locais de repouso, pequenos oásis com árvores, sombra e água.

Outra estrada Argentina, a N°40, corta o país na direção norte-sul, acompanhando a pré-cordilheira. Vai da Quebrada de Humauaca, no norte, até à Terra do Fogo, no extremo sul do país. A estrada 40 sobe e desce, atravessa vales e montanhas, uma diversidade de paisagens e terrenos.

Suponha o leitor que o Ministério de Obras Públicas tenha feito uma concorrência para a construção destas duas estradas e adjudicado os respectivos contratos a duas empresas distintas, e queira agora avaliar a eficiência de ambas com o intuito de utilizar esta informação como elemento de decisão em futuras concessões de obras públicas. Para isto, a Divisão de Informações do Ministério desenvolve indicadores de eficácia e eficiência: a quantidade de quilômetros de estrada por unidade de tempo – por exemplo, por dia ou semana – que cada empresa avançou; o custo de cada empresa por quilômetro construído de estrada. Isto feito, levanta as informações e entrega os resultados ao ministro.

Suponhamos que o ministro tome decisões com base nestas informações. Provavelmente se verá em apuros. Como é óbvio, aparentemente a empresa A, que construiu a estrada 20, será mais eficiente que a empresa B, que construiu a estrada 40. No entanto, enquanto a primeira simplesmente despejou terra, passou as máquinas niveladoras em linha reta e horizontal e colocou o material betuminoso, a segunda teve de dinamitar trechos de montanha, fazer terraplanagens, pontes para vencer cursos d’água, calcular o traçado das curvas, construir trechos em aclive e descida, etc.

<sup>17</sup> Para citar apenas um estudo abrangente de nível internacional, o TIMSS revela que *“um fator é consistente através de todas as provas do TIMSS em todas as séries: em todos os países participantes, os filhos de pais com educação universitária superaram, em média, os filhos de pais cuja educação foi interrompida no secundário, e os filhos de pais cuja educação foi interrompida no secundário, por seu lado, superaram as crianças cujos pais não concluíram o secundário. É certo também que os estudantes que têm um lar educativamente rico, com muitos livros, calculadoras e espaço para estudar, vão melhor na escola que aqueles que não o têm”* BEATON, A., 1999. Em inglês no original.

Na realidade, portanto, o ministro não sabe absolutamente nada sobre qual das duas empresas é a melhor. Pelo contrário, a informação que possui de pouco serve para esse objetivo. Suponha-se, ainda, que o ministro tenha informações sobre o tipo de tecnologia utilizada por cada empresa, com as quais conclui que a tecnologia utilizada pela empresa A é mais eficiente que aquela utilizada pela empresa B. Se todas essas análises forem feitas sem levar em conta o terreno em que cada empresa teve de trabalhar, de nada servem os indicadores desenvolvidos.

Na educação passa-se algo muito similar. A atividade educativa não se desenvolve, nem no vazio, nem com um corpo de alunos homogêneo. Os indivíduos chegam à instituição educativa em condições absolutamente diferentes para receber e aproveitar o trabalho dos educadores. Alguns tiveram na primeira infância vivida uma infinidade de estímulos visuais e afetivos, no contexto de um lar no qual rotineiramente se fala uma linguagem complexa, onde se lê habitualmente e onde o alfabeto foi ensinado antes da escola fazê-lo. Outros tiveram uma primeira infância em lares onde os estímulos visuais eram escassos, onde a linguagem usada era mais simples, onde só excepcionalmente se lia e escrevia, cujos pais mal concluíram a escola e não têm uma noção precisa do tipo de ajuda que podem dar aos filhos nos deveres de cada. Algumas escolas trabalham com alunos que se alimentam adequadamente, têm um espaço próprio em suas casas e saem de férias todos os anos. Outras trabalham com alunos que se alimentam mal, que vivem amontoados e que nunca se afastaram mais que cinco quilômetros de suas respectivas casas. Alguns estados ou províncias trabalham com uma população altamente alfabetizada e em um contexto de crescimento econômico, enquanto outros têm populações com altas taxas de adultos analfabetos e famílias em condições de extrema pobreza.

Os “terrenos” são absolutamente diferentes. Uma trabalham no plano em linha reta, e outras devem vencer uma

infinidade de dificuldades. Assim, informar diferenças entre províncias, regiões ou tipos de escola sem analisar as populações com que trabalham equivale a informar ao Ministro de Obras Públicas sobre a “eficácia” das empresas construtoras de estradas sem levar em consideração o tipo de terreno em que cada uma delas trabalhou. Poderão, pais e autoridades, tomar decisões ou formular políticas com base nos dados apresentados? Não seria conveniente advertir o usuário de que alguns professores estão construindo no plano e em linha reta, enquanto outros trabalham na pré-cordilheira?<sup>18</sup>

Qualquer análise de resultados ou indicadores educativos que não leve em conta as características do corpo de alunos tem escasso valor para a tomada de decisões, ou para extrair algum tipo de conclusão válida que pretenda ir além da mera descrição de as diferenças existem.

### III. 3. Alguns exemplos de conclusões impróprias

Transcrevem-se em seguida alguns parágrafos dos relatórios nacionais que ilustram o problema da interpretação inadequada ou excessivamente simplificada dos resultados, na medida em que induzem o leitor a tirar conclusões sobre a eficácia do ensino nos agregados descritos sem alertar para a incidência da diferenciação sociocultural das populações de tais agregados.

*“Em ambos os regimes escolares, nos três anos e em todas as destrezas avaliadas, as turmas de escolas particulares urbanas superaram em percentuais de domínio as turmas das escolas públicas urbanas, e estas últimas*

---

<sup>18</sup> Algo similar ocorre quando os economistas, como o fazem normalmente, interpretam os diferenciais de repetência e de alunos fora da faixa etária como ineficiências do sistema educativo. Na realidade, eles indicam que em certos setores da sociedade o terreno para o trabalho educativo é mais complicado, razão pela qual leva-se mais tempo para educar.

*superaram as das escolas rurais. A análise de variância demonstra que nas terceira e sétima séries, a vantagem das turmas de escolas particulares urbanas sobre as de escolas públicas urbanas, e destas sobre as das escolas rurais é estatisticamente significativa. No caso da décima série, não há diferença significativa entre as escolas públicas e rurais, mas as escolas particulares têm uma vantagem significativa sobre os dois outros tipos”* Ministerio de Educación y Cultura del Ecuador – EB/PRODEC, (1998).

*“... em qualquer das séries ou disciplinas avaliadas pelo SAEB/97, as médias de proficiência da zona urbana são sempre superiores às da zona rural. Por outro lado, os alunos da zona rural da região nordeste apresentam sempre as melhores médias (...) Os alunos da rede particular de ensino, em todas as séries, disciplinas e regiões, apresentam proficiências mais elevadas que os alunos da rede pública (estadual ou municipal)...”* Ministério da Educação - Governo Federal. INEP, 1998.

Na realidade, as afirmações transcritas nos parágrafos anteriores estão, todas, estritamente certas e se limitam a descrever dados empíricos. Apesar disto, colocadas desta forma induzem a conclusão de que as escolas privadas são melhores ou mais eficazes que as públicas urbanas, e que estas são melhores que as rurais. Ninguém deve estranhar, portanto, que a imprensa desse manchetes deste teor ao informar os resultados da avaliação nacional.

A título de ilustração de como a leitura dos dados se altera quando se introduzem considerações sobre o contexto sociocultural, na figura 10 estão incluídas várias “classificações” de resultados por tipo de escola, elaboradas a partir dos dados do Uruguai. Na coluna acinzentada está a classificação resultante de uma consideração

global dos resultados de cada agregado sem levar em conta o tipo de alunos com que trabalham. Como nos demais países da região, as escolas privadas têm resultados notoriamente superiores às escolas públicas urbanas, e estas apresentam resultados melhores que os das escolas rurais. Nas colunas seguintes apresentam-se “classificações” que resultam de comparação das escolas controlando-se o tipo de alunos com que trabalham. Na coluna correspondente a “contexto muito favorável” – vale dizer, aquelas escolas que atendem aos setores mais privilegiados da sociedade – as escolas privadas da capital do país continuam tendo melhores resultados que as públicas, mas a diferença estreitou-se, baixando de 24 pontos percentuais para apenas 5,4. As escolas privadas do interior passaram ao terceiro lugar. As escolas rurais não aparecem nesta coluna porque, como é óbvio, não há escolas rurais que atendam a uma população altamente favorecida.

Figura 10

URUGUAI. Uma “classificação” que se altera em função do contexto sociocultural das escolas. Percentuais de alunos “suficientes” em Matemática

| Posição | Global                     | Contexto Muito Favorável   | Contexto Favorável         | Contexto Desfavorável      |
|---------|----------------------------|----------------------------|----------------------------|----------------------------|
| 1º      | Privadas Capital<br>61,2%  | Privadas Capital<br>71,2 % | Públicas Int.              | Rurais<br>32,1%            |
| 2º      | Privadas Interior<br>47,1% | Públicas Capital<br>65,8%  | Públicas Capital<br>46,4%  | Privadas Interior<br>29,2% |
| 3º      | Públicas Capital<br>35,3%  | Privadas Interior<br>61,1% | Privadas Interior<br>44,4% | Públicas Int.              |
| 4º      | Públicas Int.              | Públicas Int.              | Privadas Capital<br>41,6%  | Públicas Capital<br>23,2%  |
| 5º      | Rurais<br>19,4%            |                            |                            | Privadas Capital<br>15,4%  |

Fonte: ANEP, Unidad de Medición de Resultados Educativos, Uruguay, 1997. *Evaluación Nacional de Aprendizajes en Lengua y Matemática. 6º año de Educación Primaria - 1996. Segundo Informe de Difusión de Resultados.*

Quando se passa à coluna seguinte – que já não corresponde aos setores mais favorecidos, mas aos setores de classe média ou classe média-alta – as escolas públicas apresentam melhores resultados que as privadas, tanto na capital como interior do país. Neste setor também não existem escolas rurais. Finalmente, quando se comparam os resultados entre as escolas que atendem a setores desfavorecidos da população, constata-se que, na realidade, as escolas rurais são “melhores” que as demais.

Outro exemplo ilustrativo de conclusões impróprias, no qual, além de um problema de omissão da consideração do contexto ao informar os resultados, também se formulam explicitamente conclusões a respeito da eficácia diferencial dos sistemas educativos das distintas regiões aparece no SAEB/97.

*“Os maiores percentuais de alunos de quarta série acima deste nível (175 pontos em matemática) encontram-se nas regiões Sul e Sudeste, indicando a maior eficácia do ensino nestas regiões e constituindo claro exemplo das desigualdades regionais brasileiras”.*

*“Os resultados do SAEB/97 sobre a proficiência dos alunos indicam, em síntese, a existência de dois tipos de problemas*

*relacionados com a qualidade do aprendizado em nosso país: o primeiro refere-se à grande heterogeneidade dos resultados de desempenho que existe entre os sistemas estaduais de ensino, que parece ser um fator de desigualdade mais que de simples diferenciação, na medida em que a dimensão das diferenças significa, para os alunos que apresentam desempenhos mais baixos, uma oferta de oportunidades de aprendizado menos efetiva que aquela proporcionada aos alunos que apresentam desempenhos mais altos”* Ministério da Educação – Governo Federal. INEP, 1998.

Qualquer leitor medianamente informado sabe que os estados no nordeste brasileiro são os mais pobres e menos alfabetizados. Pelo contrário, os do sul são os que têm maior desenvolvimento econômico e cultural. Portanto, mesmo que seja importante mostrar que as crianças e os jovens que vivem nos primeiros estão em clara desvantagem frente aos segundos, e que existe uma forte desigualdade, é delicado atribuir estas diferenças à eficácia dos sistemas de ensino, dos currículos ou do trabalho dos educadores.

O problema é que os diferentes dados não se conectam entre si. Por um lado analisam-se os resultados por jurisdição, área geográfica e/ou tipo de escola, e por outro a

relação dos mesmos com a origem social dos estudantes. **Falta ainda dar o passo de inter-relacionar estas informações, fazer uma análise e uma apresentação mais complexas dos dados, e alertar o usuário para essas complexidades e os cuidados na interpretação, buscar modos de estabelecer comparações entre setores do sistema educativo que trabalhem em “terrenos” similares para poder formular conclusões válidas a respeito da eficácia dos distintos tipos de escola, currículos ou sistemas de ensino.**

O tratamento dos resultados das avaliações nacionais sem uma adequada

consideração do contexto social e do tipo de alunos com que trabalham as escolas é mais grave quando as avaliações têm implicações “fortes” para as escolas, como ocorre no caso chileno. No relatório dos resultados nacionais correspondentes à 8ª série do Ensino Básico, avaliados no ano de 1995, publica-se a lista dos vinte melhores estabelecimentos do país, que são aqueles que tiveram as pontuações médias mais altas. Destes, 19 são escolas privadas, 12 estão localizados em bairros, de boa situação econômica, de Providencia e Las Condes, na capital do país (ver figura 11).

**Figura 11**  
**CHILE. Estabelecimentos com as vinte médias nacionais mais altas em matemática e castelhano**

| Nº ORDEN | DEPENDÊNCIA     | DEPTO. PROVINCIAL | COMUNA       | ESTABELECIMENTO                         |
|----------|-----------------|-------------------|--------------|---|
| 1        | Particular pago | Cautín            | Villarrica   | <b>Colegio Alemán</b>                   |
| 2        | Municipal       | Santiago Oriente  | Providencia  | <b>Liceo Carmela Carvajal de Prat</b>   |
| 3        | Particular pago | Talca             | Constitución | <b>Colegio Constitución</b>             |
| 4        | Particular pago | Santiago Oriente  | Las Condes   | <b>Anglo American International</b>     |
| 5        | Particular pago | Santiago Oriente  | Vitacura     | <b>Colegio Los Andes</b>                |
| 6        | Particular pago | Santiago Oriente  | Las Condes   | <b>Colegio Cumbres</b>                  |
| 7        | Particular pago | Santiago Oriente  | Las Condes   | <b>Colegio La Girouette</b>             |
| 8        | Particular pago | Elqui             | La Serena    | <b>Colegio La Serena</b>                |
| 9        | Particular pago | Valparaíso        | Valparaíso   | <b>Seminario San Rafael</b>             |
| 10       | Particular pago | Valparaíso        | Viña del Mar | <b>Seminario San Rafael</b>             |
| 11       | Particular pago | Santiago Oriente  | Las Condes   | <b>Colegio La Abadía</b>                |
| 12       | Particular pago | Santiago Oriente  | Las Condes   | <b>Colegio Nuestra Señora del Pilar</b> |
| 13       | Particular pago | Santiago Oriente  | Providencia  | <b>Saint Gabriel's English School</b>   |
| 14       | Particular pago | Santiago Oriente  | Las Condes   | <b>Villa María Academy</b>              |
| 15       | Particular pago | Santiago Oriente  | Las Condes   | <b>Colegio del Sagrado Corazón</b>      |
| 16       | Particular pago | Santiago Oriente  | Providencia  | <b>Saint Gabriel's English School</b>   |
| 17       | Particular pago | Santiago Oriente  | La Reina     | <b>Nuestra Señora del Camino</b>        |
| 18       | Particular pago | Santiago Oriente  | Providencia  | <b>Colegio Huelén</b>                   |
| 19       | Particular pago | Osorno            | Osorno       | <b>Instituto Alemán de Osorno</b>       |
| 20       | Particular pago | Santiago Oriente  | Providencia  | <b>Colegio St. Thomas Morus</b>         |

**Fonte:** República de Chile, Ministerio de Educación, SIMCE, 1996. *Resultados 95 - 8º año Básico.*

No mesmo relatório, elabora-se uma classificação dos estabelecimentos por “nível socioeconômico” com três categorias (A, B e C) em função do gasto mensal médio em educação e do nível de educação dos pais dos alunos. A distribuição dos tipos de escolas segundo seu nível socioeconômico mostra claramente que as escolas privadas pertencem à categoria A e B, enquanto a enorme maioria das escolas municipais pertence à categoria C (figura 12). Não obstante, apesar dessas constatações, toda a apresentação dos resultados enfoca quase que exclusivamente a comparação entre as pontuações médias globais dos estabelecimentos municipais subvencionados e dos particulares pagos.

**Figura 12**  
**CHILE. Distribuição dos estabelecimentos por nível socioeconômico segundo tipo de administração**

| Nível        | Particular Pago |              | Subvencionado |              | Municipal  |              |
|--------------|-----------------|--------------|---------------|--------------|------------|--------------|
|              | Quantidade      | Percentual   | Quantidade    | Percentual   | Quantidade | Percentual   |
| <b>A</b>     | 287             | <b>60,4</b>  | 66            | <b>4,9</b>   | 8          | <b>0,3</b>   |
| <b>B</b>     | 188             | <b>39,6</b>  | 517           | <b>38,5</b>  | 342        | <b>12,3</b>  |
| <b>C</b>     | 0               | <b>0,0</b>   | 759           | <b>56,6</b>  | 2434       | <b>87,4</b>  |
| <b>Total</b> | 475             | <b>100,0</b> | 1342          | <b>100,0</b> | 2784       | <b>100,0</b> |

**Fonte:** República de Chile, Ministerio de Educación, SIMCE, 1996. *Resultados 95 - 8º año Básico.*

No caso chileno, esta deficiência na apresentação dos resultados foi parcialmente sanada a partir da medição de 1997, quando se incorporou o dado do progresso de cada estabelecimento em relação a seu próprio resultado em 1995 (ver o capítulo IV). Isto é muito mais apropriado que a comparação com os resultados brutos de outros estabelecimentos que, em geral, trabalham com populações muito diferentes.

### **III. 4. A consideração do contexto sociocultural**

Assim como a maneira válida de elaborar indicadores comparativos da eficiência das empresas construtoras de estradas seria medir sua taxa de avanço em terrenos similares, **a única maneira válida de elaborar indicadores que sirvam para sustentar afirmativas ou decisões sobre a eficácia das instituições ou sistemas educativos é através de algum tipo de categorização dos “terrenos” em que trabalham, vale dizer, das características socioculturais dos lares de onde provêm seus alunos.**

Um exemplo que ilustra um dos caminhos possíveis para abordar este desafio é a elaboração de categorias de “contexto sociocultural das escolas”<sup>19</sup>, com a finalidade

<sup>19</sup> É importante destacar que o “**contexto sociocultural**” é uma propriedade da escola ou do grupo de alunos, não de indivíduos.

de comparar os resultados entre escolas que tenham grupos de alunos similares. A constatação de que o ordenamento dos resultados das provas nos diferentes tipos de escola segue o mesmo ordenamento que os indicadores sociais deu lugar, no Uruguai, a uma classificação dos estabelecimentos educativos em cinco categorias segundo seu

**Primordialmente, o que explica os resultados não é a origem social individual dos estudantes, mas a composição sociocultural do grupo escolar ou da escola a que pertencem.** Esta advertência é de grande importância porque em algumas análises comparam-se, por exemplo, os resultados de indivíduos de nível sociocultural “baixo” que pertencem, de um lado a escolas públicas, de outro a escolas privadas. Normalmente as diferenças não serão significativas, e se o forem, os resultados serão melhores entre os alunos das escolas privadas, mas isto como consequência do fato de que em tais escolas os indivíduos de nível socioeconômico “baixo” são a minoria, enquanto nas primeiras são maioria. Em outras palavras, quando indivíduos de origem sociocultural mais baixa são minoria em um grupo constituído majoritariamente por indivíduos de melhor “respaldo” cultural familiar, os primeiros serão favorecidos por seus colegas. Há um efeito do nível do grupo que melhora os aprendizados. Em troca, quando o grupo é majoritariamente constituído por indivíduos de origem social desfavorável, a minoria com melhor nível educativo familiar tenderá a ter resultados inferiores aos esperados. Estas constatações, por outro lado, indicam a importância de favorecer-se a heterogeneidade social nas escolas.

“contexto sociocultural”: “muito favorável”, “favorável”, “médio”, “desfavorável” e “muito desfavorável”. Esta categorização foi elaborada a partir de dois índices: um, de caráter cultural, desenvolvido a partir dos níveis educativos dos lares dos alunos; e o outro, de caráter econômico, construído a partir dos níveis de equipamentos de tais domicílios.

A partir desta elaboração, os resultados passaram a ser apresentados controlando sempre o contexto sociocultural, como na

figura 10 – vale dizer, mostrando os resultados por tipo de escola, região ou departamento dentro de uma mesma categoria de contexto sociocultural. Além disto, ao dar feedback às escolas, entrega-se a cada uma delas o resultado de seus alunos junto com o correspondente às demais escolas do mesmo contexto sociocultural, de modo que possam se comparara, não apenas com a média nacional, como também com as escolas de composição social similar. (figura 13).

**Figura 13**  
**URUGUAI. Quadro típico de apresentação dos resultados a cada escola**

|   | <b>Escola</b> | <b>Escolas do mesmo contexto sociocultural do Departamento</b> | <b>Escolas do mesmo contexto sociocultural do país</b> | <b>Total Nacional</b> |
|---|---------------|--|--|-----------------------|
| <b>Percentual de alunos suficientes em língua materna</b>                     |               |  |  |                       |
| <b>Percentual de alunos suficientes em compreensão de texto argumentativo</b> |               |  |  |                       |
| <b>Percentual de alunos suficientes em compreensão de texto narrativo</b>     |               |  |  |                       |
| <b>Percentual de alunos suficientes em reflexões sobre a língua</b>           |               |  |  |                       |

### **III. 5. Os dilemas que a relação entre aprendizados e contextos sociais coloca na divulgação dos resultados**

O tratamento e a apresentação dos resultados das avaliações nacionais a partir da categorização dos contextos sociais gera um novo problema: corre-se o risco de acentuar uma visão determinista, na sociedade e entre os educadores, visão esta que poderia ser ilustrada através de frases do tipo “isto é o máximo que se pode conseguir neste meio” ou “neste contexto não se pode esperar nem

exigir mais das crianças”. Em outras palavras, corre-se o risco de estabelecer ou reforçar um sistema de expectativas diferenciadas (“exijamos mais dos setores médios e altos; exijamos menos dos setores populares porque o que podem atingir é diferente”). Muitas pessoas vinculadas ao setor educativo e aos sistemas de avaliação temem, justificadamente, que enfatizar a apresentação de resultados diferenciados por contexto social pode operar, para os professores que trabalham nesses contextos,

ou como uma “cortada”, ou como justificativa para os resultados insuficientes.

Mas, por outro lado, ignorar o contexto pode dar lugar a conclusões espúrias e, mais ainda, pode gerar uma grande frustração e impotência nas escolas e professores que trabalham nos meios mais desfavorecidos, na medida em que pareça que toda a responsabilidade pelos péssimos resultados é da escola. Pode dar lugar, inclusive, a que certas escolas percam injustamente sua legitimação perante suas comunidades, com efeitos paralisantes. A avaliação pode ser percebida como injusta porque responsabiliza o professor e a escola por problemas que, ao menos em parte, provêm do contexto. Voltando à parábola das estradas, é como se a empresa que trabalhou na pré-cordilheira ficasse, em uma classificação por média de metros construídos por dia, abaixo da que trabalhou na planície. Sentir-se-ia fortemente injustiçada. **A introdução da noção de contexto social “alivia” o peso da responsabilidade, ajuda a perceber a avaliação como equitativa<sup>20</sup> e a compreender que há fatores externos ao sistema educativo que afetam o aprendizado, e que a educação não é onipotente. Ajuda, definitivamente, a estabelecer algo que é estritamente certo: parte das desigualdades que se registram no sistema educativo é reflexo das desigualdades que existem na sociedade.**

Algumas formas de enfrentar este dilema são:

- Mostrar de forma sistemática, através da divulgação dos resultados, que mesmo no interior dos diferentes contextos sociais há escolas que obtêm melhores resultados que outras. Em outras palavras, que, mesmo que o contexto social tenha um peso específico sobre os resultados, não há determinismos absolutos e existe um espaço próprio da ação pedagógica e escolar que é necessário fortalecer. Esta pode ser, também, uma forma de motivar para que, dentro do sistema educativo, se faça todo o possível para obter

maior equidade na distribuição social do conhecimento e para melhorar a capacidade das crianças de origem mais pobre.

- Dar forte ênfase aos dados longitudinais, e não às comparações transversais entre escolas. O esquema ideal seria fazer-se uma medição inicial e outra final em um mesmo ano letivo. Deste modo pode-se eliminar uma boa parte do efeito do contexto, concentrando a medição e o relatório de resultados no que os alunos e turmas progredem ao longo de um ano em comparação com seu próprio ponto de partida. Com bons instrumentos de medição de práticas e enfoques de ensino, uma abordagem deste tipo deveria melhorar a capacidade explicativa dos “fatores escolares” e, portanto, a acumulação de conhecimento neste terreno<sup>21</sup>.

Nos países que trabalham com avaliações censitárias, concentrar o relatório de resultados nas mudanças que se registram em cada escola com relação à medição anterior. Nesta alternativa, que se aplica no Chile, é necessário, de qualquer modo, algum controle sobre a composição social do quadro de alunos porque as mudanças nos resultados de uma escola podem ser consequência de mudanças no tipo de alunos que recruta.

#### **Capítulo IV. A Utilidade Pedagógica dos Relatórios de Resultados**

Do que foi exposto no primeiro capítulo depreende-se que um dos principais caminhos através dos quais espera-se que as informações produzidas pelos sistemas nacionais de avaliação tenham um efeito de melhorar o sistema educativo é sua utilização por parte das equipes dirigentes e docentes das escolas para tomada de decisões de

---

<sup>21</sup> Obviamente, um dispositivo deste tipo é mais oneroso e mais complexo quanto ao desenho das provas. Uma experiência deste tipo que deveria ser observada está sendo iniciada na Província de Buenos Aires, com o apoio do OREALC.

---

<sup>20</sup> Ver BENVENISTE, L., 2000.

caráter técnico-pedagógico que lhes permitam melhorar suas práticas de ensino.

A elaboração de relatórios de resultados para uso das equipes docentes tem sido um dos aspectos mais trabalhados da divulgação dos relatórios nos países da região. Argentina, Chile, Equador, Paraguai e Uruguai produziram relatórios específicos dirigidos aos professores nos quais se explicam, com certo detalhe, o que foi avaliado pelas provas aplicadas e quais foram os principais problemas e dificuldades detectados.

Quatro estratégias principais foram adotadas com esta finalidade:

- a produção e distribuição a todas as escolas de materiais e textos didáticos derivados dos resultados gerais da avaliação, fosse ela de tipo censitário ou por amostragem;
- a entrega a cada estabelecimento dos resultados de seus próprios alunos, no contexto de aplicações de caráter censitário;
- realização de operações de avaliação por amostragem com aplicação e correção autônomas no universo das escolas;
- realização de seminários de divulgação dos resultados e programas de capacitação em serviço a partir dos mesmos, dirigidos a professores e diretores e/ou supervisores.

#### **IV.1. Materiais de orientação didática elaborados a partir dos resultados**

A Argentina foi provavelmente o país que mais ênfase deu, desde o início de suas avaliações nacionais, à elaboração de materiais de análise das competências mais e menos aprendidas nas diferentes áreas avaliadas, através dos assim-chamados *cuadernos metodológicos* (cadernos metodológicos) sistematicamente publicados desde a primeira avaliação nacional, realizada em 1993.

Tais cadernos, que de certo modo serviram de modelo a outros países da região,

têm como objetivo contribuir para a reflexão didática e disciplinar dos professores a partir de exemplos concretos de questões e seus resultados, enfatizando as atividades que foram realizadas em maior proporção, bem como as que se provaram mais difíceis.

A figura 14 ilustra o tipo de análise de questões que é entregue aos professores. Como se pode ver, apresenta uma análise detalhada da operação cognitiva exigida para resolver a atividade, assim como das dificuldades que podiam levar os alunos a distintas alternativas incorretas, ou “distrações”.

Esta maneira de comunicar os resultados permite dar uma visão sobre o tipo de tarefas que foram enfrentadas pelos alunos nas provas. Além disto, o trabalho de explicitar o tipo de operação cognitiva envolvida na solução da atividade e o tipo de dificuldade que pode ter levado o aluno a escolher uma alternativa incorreta constitui informação valiosa para a análise dos resultados por parte dos professores. Ao mesmo tempo, cabe assinalar como fraqueza uma atomização da visão das competências visadas, na medida em que a análise se apóia em questões individuais que, em si mesmas, não representam necessariamente a complexidade de uma competência. Esta dificuldade está especialmente presente na área de Língua. Um caminho adequado para enfrentar-se esta fraqueza é o que foi adotado pelo SAEB e mencionado no capítulo II: informar as escalas de proficiência (matrizes de referência) que descrevem o que os alunos são capazes de fazer nas distintas faixas de pontuação das provas.

Ainda que a análise da qualidade e da base conceitual sobre a qual se formulam as recomendações em cada país esteja fora do alcance do presente trabalho, a impressão geral é que, assim como acontece com as provas, há uma forte heterogeneidade entre os países da região quanto à atualização dos enfoques disciplinares e didáticos a partir dos quais são elaborados, tanto os instrumentos de avaliação, como os cadernos de caráter didático.

Figura 14  
ARGENTINA. Análise de questões

**7º AÑO EGB / Matemática**

**18** Si al descarozar 20 kg de ciruelas se obtienen 15 kg de pulpa, ¿qué porcentaje de las ciruelas es la pulpa?

A) 5 %  
B) 25 %  
C) 35 %  
▶ D) 75 %

**Respuesta correcta: D**

---

**Resultado:** El 22% de los alumnos lo resolvió correctamente.

**Contenido:** Números y operaciones.

**Capacidad:** Resolver problemas.

**Operación requerida y evaluada:** Aplicar el concepto y cálculo de porcentaje para resolver un problema.

- ▶ La elección de la respuesta correcta **D** permite suponer que los alumnos comprendieron el problema, reconocieron los pasos a seguir, plantearon correctamente la situación y calcularon el porcentaje solicitado.
- ▶ La elección de la respuesta **NO** correcta **A** indica una incorrecta lectura del enunciado, los alumnos que la eligieron operan restando los datos, sin advertir que se trata de diferentes magnitudes.
- ▶ La elección de la respuesta **NO** correcta **B** indica que estos alumnos pueden calcular porcentaje, responden al porcentaje de carozos y no de pulpa, marcando así un error de interpretación.
- ▶ La elección de la respuesta **NO** correcta **C** señala que los alumnos parecen no dominar el concepto de porcentaje, sumando todos los datos que aparecen en el enunciado.

**Fonte:** Ministerio de Cultura y Educación de la Nación - República Argentina. Dirección Nacional de Evaluación, 2000. *Operativo Nacional de Evaluación 1999*; pág. 56.

No caso do Equador, por exemplo, a análise do erro e da dificuldade identificada é de carácter muito mais geral (figura 15). Este último exemplo é bastante representativo do

que se faz na maioria dos países ao propor exemplos de questões, que em geral são entregues com um escasso nível de conceituações didática e disciplinar.

**Figura 15**  
**EQUADOR. Análise de questões**

| Un automóvil recorre 80 km. en una hora. ¿Cuántos kilómetros avanzará en una hora y tres cuartos si mantiene la misma velocidad? |          |               |                                 |
|--|----------|---------------|---------------------------------|
|  | Opciones | % de elección | Observaciones                   |
| A.   | 160 km.  | 22.20         |                                 |
| B.   | 140 km.  | 34.73         |                                 |
| C.   | 120 km.  | 28.35         | no tienen claro lo que se pide. |
| D.   | 60 km.   | 6.94          |                                 |
| ERROR: identificación incorrecta de la información.  |          |               |                                 |
| DIFICULTAD IDENTIFICADA: insuficiente comprensión de la proporcionalidad y poca capacidad de análisis.                           |          |               |                                 |

| Un equipo de 4 personas en 8 horas tiende 320 metros de redes eléctricas. ¿Cuántas personas se debe emplear para que tiendan 2 400 metros en 12 horas? |             |               |  |
|--|-------------|---------------|--|
|  | Opciones    | % de elección | Observaciones  |
| A.   | 45 personas | 12.89         |  |
| B.   | 30 personas | 14.43         |  |
| C.   | 20 personas | 40.66         |  |
| D.   | 6 personas  | 21.09         | consideran los datos de personas y horas y no retienen los referidos a metros. |
| ERROR: retención parcial de la información.  |             |               |  |
| DIFICULTAD IDENTIFICADA: poca capacidad de análisis.   |             |               |  |

**Fonte:** Ministerio de Educación y Cultura - Ecuador. EB/PRODEC - Sistema Nacional de Evaluación de la Educación APRENDO, 1997. *Análisis de las Pruebas Aprendo 1996 y de sus Resultados*; pág. 25.

#### IV. 2. Entrega de resultados por estabelecimento

A entrega a cada estabelecimento de seus próprios resultados, no contexto de avaliações nacionais de caráter censitário, foi realizada na Argentina, no Chile e no Uruguai, havendo também experiências em nível subnacional, como o caso do estado do Paraná, no Brasil, que desenvolveu seu próprio sistema estadual de avaliação com esta finalidade, ou de algumas províncias argentinas, como Mendoza. Ambas são avaliações de caráter censitário, e por isto todos os estabelecimentos educativos do estado e da província recebem seus resultados<sup>22</sup>.

A principal vantagem de entregar resultados aos estabelecimentos é que isto pode envolvê-los diretamente no processo de avaliação. Há grande probabilidade de que a predisposição a utilizar os resultados seja maior quando o estabelecimento participou da avaliação e tem diante de si os resultados de seus próprios alunos. Pelo contrário, é provável que esta predisposição seja muito menor quando se recebe, simplesmente, um relatório de resultados nacionais, sem um ponto de referência sobre até que ponto os próprios alunos dominam os aprendizados avaliados.

No caso do Estado do Paraná, no Brasil merece especial destaque a pauta que se entrega a cada estabelecimento para que a equipe docente elabore um relatório de interpretação dos resultados (figura 16). A estratégia é interessante, porque de certo modo “obriga” todas as equipes a analisar seus respectivos resultados e a fazer uma

interpretação por escrito sobre os mesmos e elaborar um plano de melhoria.

Um matiz importante entre os países quanto à estratégia de devolução dos resultados aos corpos docentes é o que se refere ao acesso às provas aplicadas. Na maior parte dos países as provas não têm caráter público, informando-se unicamente sobre os conteúdos, competências ou objetivos que foram avaliados e divulgando-se exemplos do tipo de questões aplicado.

Nos casos do Uruguai e do Paraná as provas são públicas (enquanto os resultados dos estabelecimentos individuais são confidenciais, como indicado no capítulo I). Cada escola recebe exemplares das provas e manuais de interpretação das atividades incluídas nas mesmas – com uma estrutura muito similar àquela ilustrada nos casos da Argentina e do Equador, mas abrangendo a totalidade dos itens de cada prova – bem como os resultados dos alunos do estabelecimento em cada uma de tais atividades. No caso do Chile, a partir de setembro de 1999, o *Ministerio de Educación* começou a tornar públicas as provas que haviam sido aplicadas em 1998<sup>23</sup>.

Um aspecto interessante do caso chileno, mencionado no capítulo anterior, é a apresentação sistemática de informações sobre o progresso de cada estabelecimento em relação à medição anterior. Estas informações são divulgadas, tanto através da imprensa, como da Internet.

---

<sup>22</sup> A devolução de resultados personalizados por estabelecimento também foi realizada em alguns países, como Equador ou Bolívia, cujas avaliações têm caráter de amostragem. Neste caso, a proporção de estabelecimentos que pode receber seus resultados é bastante reduzida, e a estratégia perde boa parte de seu potencial para influenciar o sistema.

---

<sup>23</sup> De todo modo, é preciso reconhecer que a divulgação das provas traz em si dificuldades na hora de elaborar novas provas para medições comparáveis que se deseje fazer no futuro.

**Figura 16**  
**ESTADO DO PARANÁ/BRASIL. Guia para análise dos resultados nos estabelecimentos educativos**

- *“Examine todas as perguntas relativas a cada conteúdo. Geralmente são três perguntas sobre cada um dos temas ou um pouco mais. Veja o grau de facilidade de cada uma delas e observe a dificuldade. Se 60% dos alunos tiverem acertado, trata-se de uma pergunta de dificuldade média, segundo a observação ao final da tabela.*
- *Examine cada item, exemplificando com erros típicos que foram observados nas provas dos alunos de sua escola para cada conjunto de questões relativas a um conteúdo. Faça comentários sobre cada conjunto e tente explicar as situações que foram identificadas.*
- *Analise, comparativamente, o desempenho dos alunos em cada um dos conteúdos, levando em consideração os dados do estabelecimento, do município, da região e do estado.*
- *A partir das questões mais difíceis observadas em sua escola, veja como ocorre a distribuição por município, região e estado. Faça uma reflexão crítica e aponte os problemas revelados através do desempenho dos alunos na prova que exigiram uma maior intervenção dos professores.*

**CONCLUSÕES DO RELATÓRIO**

*Agora você dispõe de elementos para apresentar suas conclusões no Relatório da Prova de Conteúdos Básicos. Reúna-se com seus colegas que fizeram o Relatório e elabore, em grupo, as conclusões do mesmo.*

- *Faça um comentário geral sobre as expectativas da escola em relação ao desempenho dos alunos.*
- *Manifeste-se criticamente em relação às características da prova, apresentando sugestões sobre possíveis aspectos do currículo. Trace um paralelo entre o que se pretende medir e o que a escola realmente abordou no processo de ensino nas turmas avaliadas.*
- *Apresente os pontos críticos revelados pelo desempenho dos alunos nas provas.*
- *Apóie-se em suas observações anteriores, mas procure não ser repetitivo.*
- *Dê a opinião consensual da escola sobre a avaliação, o desempenho dos alunos e a programação para os ensinos Fundamental e Médio, apresentando sugestões, levando em conta sua experiência como educador(a) e sua vivência da presente avaliação.*
- *Comente outros aspectos que julgar necessário com base nas informações que esta avaliação lhe proporcionou.*

**UTILIZAÇÃO DOS RESULTADOS NA ESCOLA**

*Faça uso do relatório e dos resultados apresentados para planejamento de suas atividades, com vistas a melhorar os procedimentos didático-pedagógicos na orientação do aprendizado.*

*Procure divulgar o relatório final da Escola entre os demais professores e entre os membros do Colegiado e da Comunidade.*

**Fonte:** Governo do Estado do Paraná, 1998. Avaliação do Rendimento Escolar. Resultados da Avaliação do Rendimento Escolar. 4ta. Série do Ensino Fundamental.

### **IV.3. Avaliações por amostragem com aplicação e correção autônomas das provas no universo das escolas**

Um terceira estratégia desenvolvida na região com o intuito de garantir que as avaliações afetem positivamente o trabalho docente consiste em:

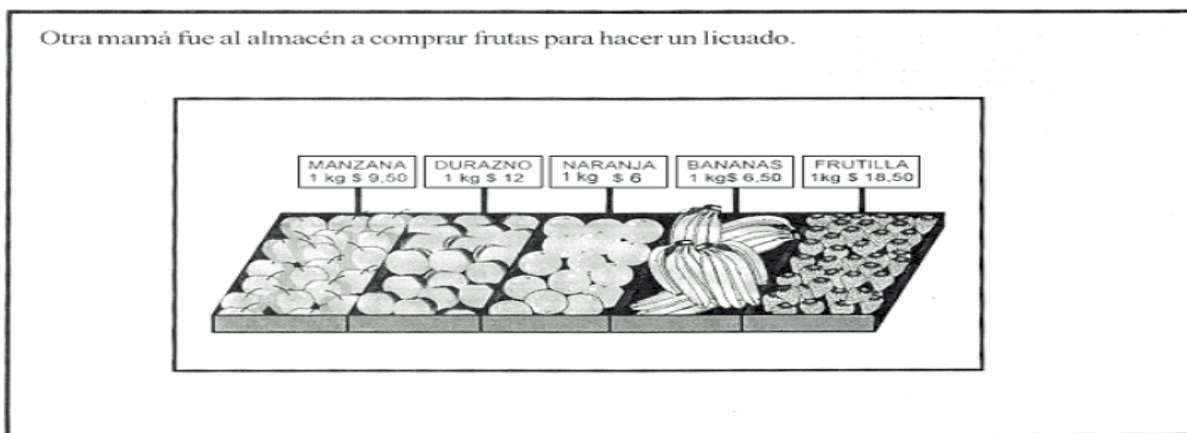
- fazer avaliação de uma amostra representativa em nível nacional;
- distribuir em seguida os instrumentos de avaliação a todas as escolas do país, junto com manuais de aplicação e correção das provas, com orientações para que tais tarefas fossem realizadas de forma voluntária e autônoma nas escolas;
- distribuir finalmente os resultados da amostra nacional para que cada escola pudesse comparar seus próprios resultados com os nacionais.

Este tipo de estratégia foi desenvolvido no Uruguai na 3ª série do primário em 1998, com uma prova de resposta aberta que integrava em um único instrumento conteúdos de Língua, Matemática e Ciências, e repetida em 1999 com provas de múltipla escolha em Língua e Matemática da 6ª série do primário.

Na figura 17 apresenta-se um exemplo dos modelos de correção das provas de resposta aberta aplicadas em 1998. Para cada uma das atividades, indicava-se aos professores o processo de resolução da atividade, uma codificação e sistema de pontos para os distintos tipos de respostas possíveis, bem como exemplos de respostas reais das crianças, tomados da etapa de aplicação piloto dos instrumentos. Além dos pontos de corte – para que cada professor pudesse estabelecer os níveis de suficiência em sua turma –, o manual incluía uma coluna onde os quadros dos resultados nacionais ficavam em branco para que cada escola pudesse ali consignar os resultados correspondentes a seus alunos. Também dava informações sobre os resultados nacionais em cada uma das atividades da prova.

O principal ponto forte desta estratégia é que envolve fortemente os professores, dando-lhes participação direta na avaliação e propiciando o desenvolvimento de uma cultura de avaliação do aprendizado com instrumentos validados sistematicamente. Seus pontos fracos são que não há garantia de que todas as escolas realmente se envolvam na avaliação e que com este tipo de instrumentos e procedimentos fica praticamente impossível realizar medições comparáveis no tempo.

**Figura 17**  
**URUGUAI - UM EXEMPLO DE PAUTAS PARA A CORREÇÃO AUTÔNOMA DE PROVAS DE RESPOSTA ABERTA NA 3ª SÉRIE DO PRIMÁRIO**



**COMPETENCIA** - Resolución de problemas.

2. Compró 3 kg. de frutas diferentes (1kg. de cada una). Gastó menos de \$25.  
 ¿Qué frutas compró?

**OBJETIVO** - Encontrar como mínimo una respuesta entre dos posibles.

**PROCESO POSIBLE PARA LLEGAR A UNA RESPUESTA CORRECTA.**

- Reconoce la suma como estrategia para relacionar los números de los datos dados.
- Suma los tres precios más bajos.
- Elige una opción cuyo resultado da una cantidad menor a \$25.
- Relaciona los tres precios sumados en dicha suma con el nombre de las frutas.

**TABLA DE VALORACIÓN**

| CÓDIGO | TIPO DE RESPUESTA  | PUNTAJE |
|--------|--|---------|
| 1      | Durazno, naranja, banana (\$24,5).<br>Manzana, naranja, banana (\$22).                           | 3       |
| 2      | Nombra 2 frutas, pero respeta el límite de gasto.  | 1       |
| 3      | Nombra 3 kilos de una sola fruta (naranja o banana) respetando el límite de gasto.               | 1       |
| 4      | a) Nombra 3 frutas, pero excede los \$25.<br>b) 3 kilos de una sola fruta, pero excede los \$25. | 0       |
| 5      | Nombra más de 3 frutas.  | 0       |
| 8      | Respuesta inadecuada.  | 0       |
| 9      | Respuesta ilegible.  | 0       |
| 0      | Ausencia de respuesta  | 0       |

**Ejemplos de respuestas dadas por los niños:**

**Código 2** 'no puede comprar 3 kg de frutillas, puede comprar naranjas y duraznos'

**Código 3** '6 + 6 + 6 = 18'

**Código 4** 'frutilla, durazno, manzana'

**Código 5** 'banana, frutilla, durazno, naranja, manzana'

**Código 5** 'naranja, banana, frutilla, durazno, manzana, uva, ponelo, zandía, mandarina'

**Código 8** '19,50 + 12 = 3070    15 + 16,50 = 5700    16,50 - 1850 = 100'

**Código 8** '\$34'

**Fonte:** URUGUAY-ANEP, Unidad de Medición de Resultados Educativos, 1998. *Evaluación Nacional de Aprendizajes en. 3er. año de Educación Primaria-1998. Manual de Corrección de la Prueba y Procesamiento del Cuestionario para madres, padres o tutores;* pág. 29 - 31.

#### IV.4. Seminários de difusão de resultados e programas de capacitação em serviço

Um quarto tipo de estratégia de divulgação de resultados entre as escolas e os professores tem sido a realização de atividades de capacitação ou sensibilização através de seminários e cursinhos, em geral dirigidos a supervisores, diretores e/ou professores e mestres. Não há, no entanto, informações sistemáticas, nem sobre as características destas ações, nem sobre seus conteúdos e cobertura: esta análise não fez parte dos objetivos deste trabalho. Não obstante, vale a pena mencionar que é provável que a distribuição de relatórios escritos não seja uma via suficiente e eficaz para garantir o aproveitamento das informações pelas equipes docentes. O desenvolvimento de instâncias, com presença física ou à distância (aproveitando as possibilidades das novas tecnologias) nas quais os resultados sejam explicados diretamente e durante as quais se façam exercícios de compreensão e interpretação dos mesmos, além da apresentação de análises de especialistas em didática e nas disciplinas avaliadas, sem dúvida amplia enormemente o aproveitamento das informações pelas escolas.

No futuro, provavelmente, as Unidades de Avaliação deverão incorporar este tipo de programas em sua missão institucional, em vez de concebê-la unicamente em termos de produção de informações válidas e confiáveis que outros saberão utilizar. Este aspecto é relevante, porque normalmente a maior parte da energia de uma Unidade de Avaliação é destinada à aplicação das avaliações, ao procedimento de informação e à produção de relatórios. Supõe-se que outras Unidades dentro dos Ministérios da Educação tenham como missão específica a organização de programas de formação ou capacitação em serviço. Sem embargo, **nem sempre existem as instâncias de coordenação e trabalho em conjunto para que os resultados das avaliações sejam incorporados aos programas de capacitação. Com isto não se pretende dizer que as Unidades de**

**Avaliação devam assumir esta tarefa – ainda que isto seja uma possibilidade e que algumas já o tenham feito –, mas que devem preocupar-se em estabelecer negociações ou estratégias deliberadas para que os resultados das avaliações sejam, de alguma maneira, incorporados aos programas de formação ou capacitação no trabalho.**

Por outro lado, uma tarefa que deveria ser assumida é a sistematização das experiências de capacitação realizadas na região a partir dos resultados das avaliações e da identificação das modalidades de trabalho mais significativas e relevantes para os professores.

#### IV. 5. Como usar as avaliações para melhorar as práticas de ensino?

Ao longo deste capítulo apresentamos as distintas modalidades ou estratégias usadas na região para garantir que os resultados das avaliações tenham algum impacto na melhoria das práticas de ensino nas escolas. Do todo modo, sobre este ponto há interrogações que continuam sem respostas pelo momento:

- Que proporção dos professores de cada país viu os relatórios de resultados e outras publicações derivadas, e que proporção os leu?
- Quanto dos relatórios foi compreendido por quem os leu?
- Quão úteis ou enriquecedoras foram as informações neles contidas?
- Os professores pedem outras coisas às Unidades de Avaliação?

Para que se possa responder a estas perguntas é necessário que se realizem estudos de caso que envolverão, tanto trabalho de pesquisas, como pesquisas de caráter qualitativo<sup>24</sup>.

---

<sup>24</sup> Sobre este tema, em princípio, o único trabalho da região foi aquele realizado por Luis Benveniste, que estudou com propósito distinto os

De toda maneira, é oportuno colocar algumas hipóteses de como tornar os resultados mais significativos em termos de melhoria das práticas de ensino nas escolas:

- é sumamente importante que os professores conheçam detalhadamente, não só a definição “formal” das competências avaliadas, como também o tipo de atividades que os alunos deverão realizar;
- é imprescindível avançar em direção a relatórios mais elaborados em termos de interpretação didática dos resultados, evitando-se os relatórios voltados exclusivamente para os resultados numéricos;
- é preciso avançar na incorporação de descrições etnográficas, derivadas de trabalhos de pesquisa de tipo cognitivo – que fariam parte do processo de desenho e administração das provas – sobre os processos reais de solução das atividades e problemas por parte de diferentes tipos de alunos, bem como sobre o tipo de dificuldades e erros que efetivamente se verificam;
- para que os resultados sejam mais significativos, é importante que os professores tenham informações sobre o desempenho de seus próprios alunos nas provas, o que pode ser feito, tanto pela via das aplicações de caráter censitário, como da aplicação autônoma das mesmas provas ou de provas equivalentes;
- Outra linha de trabalho promissora para que os sistemas de avaliação tenham impacto direto na melhoria das práticas de ensino consistira em dotar os professores de instrumentos de avaliação sistematicamente elaborados que eles possam empregar de maneira autônoma,

---

sistemas nacionais do Chile, da Argentina e do Uruguai, mas que incluiu uma série de visitas a escolas e entrevistas em profundidade com professores, diretores e supervisores a respeito da utilidade em geral dos sistemas de avaliação, não especificamente dos relatórios (ver Benveniste, L., 2000).

de acordo com suas necessidades e seu critério profissional. É provável que esta seja uma das demandas dos professores às Unidades de Avaliação, dado que a elaboração de instrumentos de avaliação de aprendizados é tarefa complexa e exige uma enorme dedicação de tempo. Talvez uma das principais contribuições que as Unidades de Avaliação poderiam dar para o fortalecimento de uma cultura de avaliação nas escolas seja colocar à disposição dos professores variadas alternativas de atividades de avaliação para conhecimentos e competências específicas, incluindo informações sobre o desempenho que os estudantes de diversos níveis e setores do sistema educativo tiveram em cada uma de tais atividades.

#### **Capítulo V. Os Relatórios de Resultados Como Contribuição Para a Formulação de Políticas Educativas**

No capítulo anterior foram analisadas as estratégias desenvolvidas para que os resultados das avaliações nacionais tenham algum impacto nas práticas de ensino nas escolas. **Neste capítulo, o enfoque é outra das finalidades explicitamente enfatizadas nas definições de objetivos dos sistemas nacionais de avaliação: a contribuição de informações relevantes e oportunas para a tomada de decisões e a formulação de políticas educativas.**

Supõe que um componente central da missão das Unidades de Avaliação seja o de dar informações sobre o estado do aprendizado – assim como sobre os fatores ou variáveis que incidem em tais Aprendizados – a diferentes “tomadores de decisão” em distintos lugares dos sistemas educativos: responsáveis pelo desenho dos currículos dos níveis central e regional; unidades encarregadas da produção de materiais educativos; instituições formadoras de professores; responsáveis pela tomada de

decisões em matéria de investimentos e pelo desenho de programas e políticas.

A este respeito, a primeira coisa a destacar é que os sistemas de avaliação dão uma contribuição central e significativa para a tomada de decisões e a formulação de políticas educativas pelo simples fato de produzirem informações sobre o que está ocorrendo com o aprendizado dos alunos. Esta função é transcendente, porque coloca no centro da questão os resultados educativos, levando-os à agenda pública. **Na falta destas informações, as discussões em torno de políticas educativas correm riscos como o de focar exclusivamente a avaliação do número de matrículas e do acesso ao sistema, de tornarem-se lutas por espaços de poder e interesses de tipo corporativo ou debates de caráter puramente teórico.** Tudo isto, como é óbvio, continua acontecendo mesmo diante de informações sobre o aprendizado. Mas, pelo menos, a existência de tais informações permite tentar focalizar as discussões na razão de ser dos sistemas educativos: garantir a todas as crianças o acesso a determinados conhecimentos e capacidades que são indispensáveis para seu desenvolvimento individual e social.

Pois bem, a partir da leitura dos relatórios nacionais, é possível identificar duas fraquezas principais sobre as quais seria preciso trabalhar mais para que as informações produzidas tenham maior impacto nas decisões de política educativa:

- o caráter excessivamente descritivo e escassamente “conceitualizado” dos relatórios;
- a visão demasiadamente simplificada sobre o estudo dos fatores escolares associados ao aprendizado e sua incidência na tomada de decisões.

### V.1. Análise, interpretação e conceituação insuficientes

A maioria dos relatórios nacionais se caracteriza por ser extremamente descritiva.

Em geral os relatórios limitam-se a relatar os dados que emergem dos primeiros processamentos, praticamente sem qualquer análise das conexões com outros dados, ou recolhidos no próprio aplicativo, ou provenientes de outras pesquisas relevantes. Tampouco fazem-se referências ou interpretações conceituais a partir da literatura existente em torno dos aprendizados avaliados, ou dos debates didáticos relacionados com seu ensino ou com os fatores que incidem sobre ele.

Como exemplo pode-se citar o relatório nacional da Costa Rica sobre a avaliação realizada em 1997. Este relatório constitui-se basicamente de 50 quadros, cada um acompanhado por umas poucas linhas que indicam o dado que mais sobressai, e finaliza com duas páginas de conclusões muito genéricas – com comentários sobre os principais dados, do tipo: *“O rendimento acadêmico nacional é mais alto na terceira série que na sexta (...) A região central supera as demais regiões no número de objetivos dominados na terceira e na sexta séries (...) O setor privado supera o setor público em ambos os níveis. Estas diferenças são estatisticamente significativas (...)”*<sup>25</sup>.

No caso da Argentina, é sintomático o fato de que seus relatórios praticamente não tenham texto e se limitem a apresentar os dados. Até a denominação do relatório é sintomática: *“Operativo Nacional de Evaluación 1999”* (Processo Nacional de Avaliação 1999). Tratam-se de relatórios sobre os

---

<sup>25</sup> Ministerio de Educación Pública, Programa de Pruebas Nacionales. Informe nacional sobre los resultados de las pruebas de diagnóstico de conocimientos 1997. Costa Rica, 1998. A afirmação de que o rendimento é mais alto na 3ª que na 6ª série refere-se a que, na primeira, os alunos conseguiram níveis de domínio dos objetivos avaliados em maior proporção que os alunos da 6ª série. Mesmo que não se trate dos mesmos objetivos, dado o enfoque adotado na avaliação da Costa Rica, é válido afirmar que haja maior ou menor nível de aprendizado dos objetivos esperados para cada curso.

aplicativos que descrevem basicamente quantas provas foram aplicadas e quais foram os principais resultados, ainda que depois sejam emitidos relatórios especificamente dirigidos aos educadores com mais detalhes sobre as competências avaliadas e as questões que apresentaram maior dificuldade.

Em alguns casos, é sabido que as conclusões são as mesmas para todos os relatórios, anos após ano. Isto significa que, na realidade, houve pouco trabalho de interpretação e análise dos dados. Esta falta

de análise fica evidente nos relatórios do APRENDO, no Equador, nos quais as mesmas conclusões repetem-se, quase que textualmente, durante três anos (ver quadro 2). Isto provavelmente é consequência, entre outras coisas, de que em muitos países – como no caso do Equador – as Unidades de Avaliação assumiram uma dinâmica excessivamente ambiciosa, de realização de aplicativos todos os anos em diversas séries e em um número cada vez maior de áreas do currículo, razão pela qual falta-lhes o tempo indispensável para uma análise em profundidade das informações produzidas.

**Quadro 2**  
**EQUADOR. Repetição de conclusões em relatórios de anos sucessivos**

*APRENDO 96: “Destaca-se o quadro de rendimento em Matemática, no qual se vêem dois pontos críticos: que nas três séries não se observa um domínio médio de nenhuma das destrezas medidas, e que as destrezas mais dominadas revelam que os aprendizados dos estudantes se concentram no nível basicamente operatório, enquanto que as destrezas menos atingidas indicam que a resolução de problemas é o aprendizado menos consolidado e amplo. Isto pode significar que a população escolar equatoriana estaria limitada quanto ao desenvolvimento do pensamento abstrato e do raciocínio lógico”. Ministerio de Educación y Cultura del Ecuador – EB/PRODEC, 1997.*

*APRENDO 97: “Destaca-se o quadro de rendimento em Matemática, no qual há dois pontos críticos: que o percentual de alunos que dominam as destrezas em cada série são notoriamente baixos e que as destrezas mais atingidas revelam que o aprendizado estudantil se concentra no nível basicamente operatório, enquanto que o aprendizado menos consolidado e amplo é aquele relativo à resolução de problemas. Isto pode significar que a população escolar equatoriana está limitada no desenvolvimento de sua capacidade de análise, que é uma expressão do pensamento abstrato e do raciocínio lógico”. Ministerio de Educación y Cultura del Ecuador – EB/PRODEC, 1998.*

*APRENDO 98: “Destaca-se o quadro de rendimento em Matemática, pois nele vêem-se dois pontos críticos: que nas três séries não se observa um domínio médio das destrezas medidas, e que as destrezas mais atingidas revelam que os aprendizados dos estudantes concentram-se no nível basicamente operatório, enquanto que as destrezas menos atingidas indicam que a resolução de problemas é o aprendizado menos consolidado e abrangente. Isto pode significar que a população escolar das CEM esteja limitada no desenvolvimento de sua capacidade de análise, que é uma expressão do pensamento abstrato e do raciocínio lógico”.. Ministerio de Educación y Cultura del Ecuador – EB/PRODEC, 1999.*

Os exemplos mencionados mostram que, em geral, as Unidades de Avaliação da região conceberam seu trabalho, até agora, principalmente em termos de produção de informação, supondo que outros saberão utilizá-las e empregá-las na tomada de

decisões. Por isto, maioria dos relatórios limita-se a relatar os dados obtidos.

Um modelo diferente e interessante de relatório de resultados foi adotado recentemente pelo Peru, país no qual durante

muitos anos o *Ministerio de Educación* não autorizou a publicação dos resultados. A partir do ano 2000, no entanto, a *Unidad de Medición de Calidad* começou a publicar os boletins denominados *Creceer*. Em cada número, os boletins tomaram um tema central baseado nos questionários que se aplicam aos alunos e professores junto com as provas. Por exemplo, a atitude dos alunos em relação às aulas de Língua e Matemática, ou os níveis de formação dos professores. Em cada boletim se faz uma revisão da literatura existente sobre o tema abordado, explica-se por que é relevante, o que revelam as pesquisas sobre o assunto particular e apresenta-se em forma de gráficos os principais dados levantados na avaliação. Do mesmo modo, os boletins destinados a informar os resultados, explicam de forma adequada o que significam as pontuações empregadas e que tipo de conclusões válidas se podem tirar dos dados a partir da metodologia adotada no desenho das provas. Trata-se de boletins breves, de boa apresentação gráfica, que tem um caráter apropriadamente reflexivo, e não meramente de relato de dados<sup>26</sup>.

Obviamente, é discutível se a função de análise das informações cabe aos sistemas de avaliação ou se estes devem limitar-se, como em geral vem ocorrendo até agora, à produção de dados. Assumindo que deveriam entregar informações mais elaboradas, um dos desafios centrais para os próximos anos seria incorporar a função de “analistas de informações” de diversas especialidades – didática das distintas disciplinas, ciências sociais, políticas educativas, para mencionar apenas algumas – que trabalhem mais os dados e os interpretem, formulando hipóteses, recomendações e pesquisas.

## **V.2. Os “fatores escolares” associados aos aprendizados e às políticas educativas**

A insuficiência das análises, interpretações e conceituações dos dados fica evidente no tratamento que se deu, em muitos países, à questão dos assim-chamados “fatores escolares associados” aos resultados. Nos relatórios nacionais detectam-se dois problemas principais:

- uma deficiência metodológica que consiste no fato de que alguns relatórios se apóiam exclusivamente em uma análise bivariada;
- uma visão de certo modo ingênua sobre a relação entre a pesquisa sobre fatores associados e a tomada de decisões de política educativa.

Em relação ao primeiro, cabe assinalar que é sumamente arriscado, para dizer o mínimo, formular algum tipo de conclusão em termos de pesquisa ou de recomendação de política educativa a partir de análise de tipo bivariado entre variáveis escolares e resultados. Para que se possa fazê-lo, é necessário, em primeiro lugar, ou controlar o efeito da composição social do grupo, ou trabalhar com duas medições para a mesma população, de modo que se possa medir o aprendizado de modo apropriado e neutralizar em boa parte o efeito das desigualdades de origem social dos alunos. Este aspecto foi tratado no capítulo III. Em segundo lugar, é necessário controlar a associação que os “fatores escolares” têm entre si.

Para ilustrar este último problema é útil recorrer ao Relatório do SAEB/97. Por um lado, este reconhece explicitamente a existência deste tipo de problemas e os cuidados a tomar para não simplificar a interpretação dos dados:

*“Para fins da discussão dos resultados e de sua apresentação gráfica, foram selecionadas algumas variáveis dos questionários respondidos pelos alunos e professores no SAEB/97. A seleção destas variáveis foi orientada por critérios de taxa de resposta e das possibilidades de estabelecer relações com a proficiência e o potencial de contribuição para a*

<sup>26</sup> Os boletins *Creceer* podem ser encontrados em: [www.minedu.gob.pe/web/el\\_ministerio](http://www.minedu.gob.pe/web/el_ministerio).

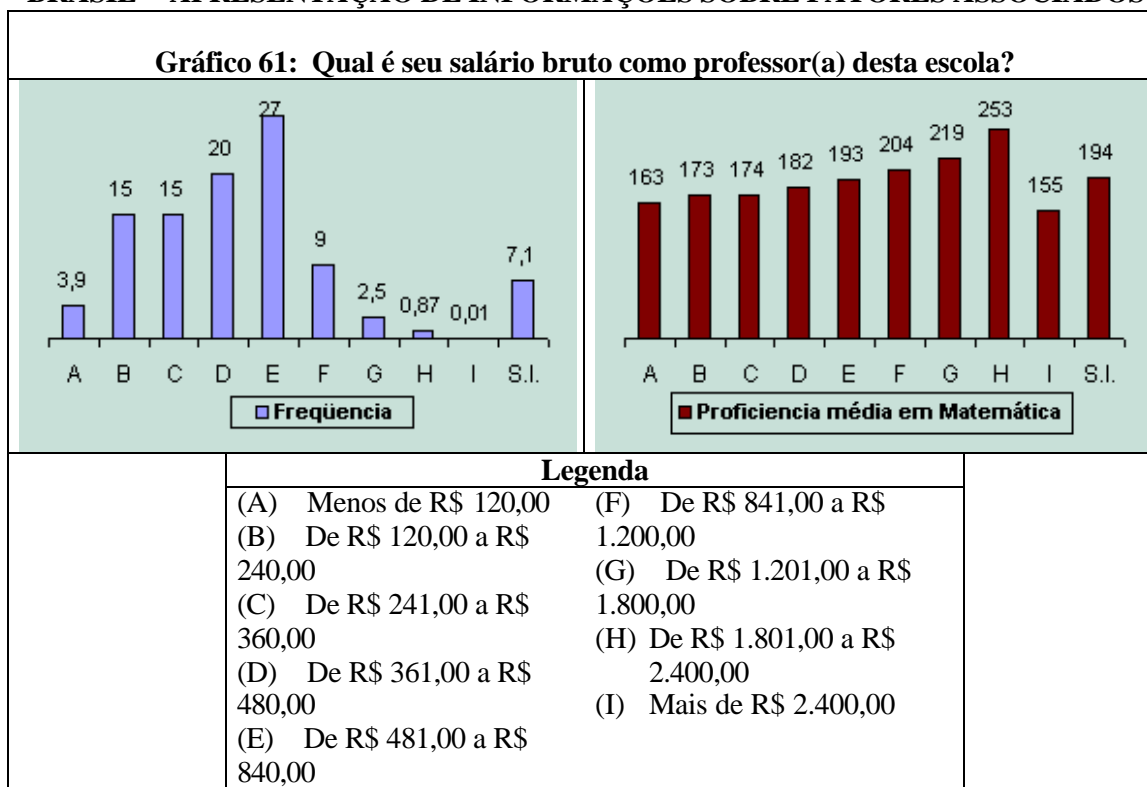
discussão de políticas e estratégias de intervenção na realidade escolar.

*É importante chamar a atenção para que a leitura dos resultados não leve a interpretações simplistas ou inferências de causa e efeito entre uma variável considerada isoladamente e o valor que lhe corresponde. A complexidade do fenômeno de aprendizado e sua múltipla causalidade devem estar sempre presentes, evitando-se reducionismos que conduziram a interpretações errôneas de certos dados”. (SAEB 1997 - Primeiros Resultados).*

Estas advertências são absolutamente pertinentes e necessárias. No entanto, o relatório inclui, na seqüência, um conjunto de quadros bivariados que conduzem precisamente aos erros de interpretação a respeito dos quais se havia alertado (ver figuras 18 e 19). Qualquer leitor médio não especializado, mesmo que tenha sido previamente advertido, provavelmente concluirá que: a) a melhores salários docentes, correspondem melhores resultados em Matemática; e b) a participação em cursos de atualização não tem efeito na melhoria dos aprendizados (pelo contrário, piora-os).

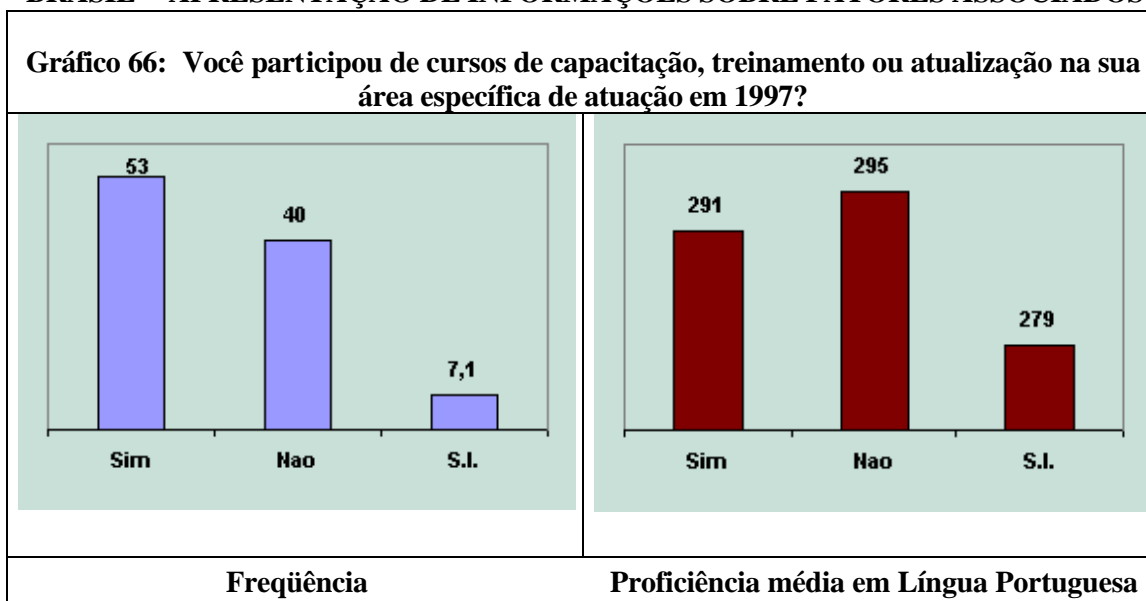
**Figura 18**

**BRASIL - APRESENTAÇÃO DE INFORMAÇÕES SOBRE FATORES ASSOCIADOS**



**Fonte:** Ministério da Educação - Governo Federal. Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), 1998. *SAEB 97 - Primeiros Resultados*.

**Figura 19**  
**BRASIL - APRESENTAÇÃO DE INFORMAÇÕES SOBRE FATORES ASSOCIADOS**



**Fonte:** Ministério da Educação – Governo Federal. Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), 1998. *SAEB 97 - Primeiros Resultados*.

Pois bem, os dados incluídos nas figuras 18 e 19 podem ser objeto de múltiplas explicações ou interpretações. Provavelmente a associação entre salários e resultados deve-se ao fato de que os salários mais altos correspondem às escolas privadas, mais caras, e, portanto, a alunos provenientes de setores médios e altos da sociedade. Do mesmo modo, a participação em cursos de capacitação nada diz sobre a heterogeneidade e a qualidade dos mesmos. Também poderia ocorrer que aqueles que participam mais em tais cursos sejam os professores mais jovens e de pouca experiência, explicando-se assim os resultados inferiores de seus alunos. Enfim, dificilmente se poderá chegar a algum tipo de conclusão válida a partir da análise bivariada.

Apesar disto, o relatório do SAEB, referindo-se ao gráfico 66, afirma:

*“Enquanto 53% dos professores declaram ter participado de cursos de treinamento, capacitação e atualização, 40% não realizaram nenhuma dessas atividades. É interessante notar que esta variável parece não ter nenhuma influência no nível de proficiência dos alunos”* (o destaque é nosso). Mais adiante,

o relatório conclui com uma afirmação do seguinte teor: *“Neste sentido, o exame dos resultados de proficiência associados a algumas características dos alunos e dos professores pode dar importantes pistas sobre quais possam ser os novos direcionamentos das práticas de alunos, professores, escolas e famílias. Os dados do Sistema Nacional de Avaliação da Educação Básica são uma preciosa fonte de estudos e, principalmente, garantem o monitoramento de aspectos relacionados com a oferta educacional e com os resultados que vêm sendo obtidos pelo sistema educacional brasileiro, principalmente aqueles relacionados com a equidade”*<sup>27</sup>.

Novamente, o que foi dito está estritamente correto ao afirmar que os dados constituem uma preciosa fonte de estudos, mas de outro tipo de estudos mais

<sup>27</sup> Ministério da Educação - Governo Federal. Instituto Nacional de Estudos e Pesquisas Educacionais (INEP). SAEB 97 - Primeiros Resultados.

sofisticados que, por outro lado, o próprio SAEB desenvolveu posteriormente<sup>28</sup>.

O segundo aspecto que é preciso mencionar é a visão um tanto simplista e ingênua da relação direta entre dados e decisões ou políticas, sem outro tipo de mediações. O tema foi analisado em uma publicação anterior do PREAL<sup>29</sup>. Muitas vezes os relatórios nacionais dão a impressão de que se espera que, a partir da pesquisa sobre os “fatores associados”, se possam formular recomendações específicas de política educativa. Inclusive em relatórios produzidos por organismos internacionais de crédito era possível encontrar, alguns anos atrás, recomendações simplistas do tipo “é preciso investir em livros-textos, mas não em formação de professores”, ou “a quantidade de alunos por turma não é uma variável relevante para a melhoria do aprendizado”, formuladas a partir de revisões de literatura baseadas em análises excessivamente simplificadas de “fatores associados”.

Sobre este tema, o caso do Equador é interessante e ilustrativo. Através de uma análise de tipo bivariado, o relatório conclui que é difícil estabelecer políticas homogêneas para todo o sistema educativo a partir dos dados analisados, porque os fatores associados aos aprendizados variam entre os diferentes contextos. O tratamento que se dá no texto a esta constatação parece indicar algum tipo de “culpa” por não se ter chegado a conclusões mais contundentes para a tomada de decisões (ver quadro 3). No entanto, o fato de que se tenha chegado a esta constatação é sumamente relevante e o relatório deveria tê-lo ressaltado de maneira mais enfática para as autoridades, alertando sobre as falsas expectativas ou visões simplistas sobre a política educativa. A constatação de que os fatores que incidem

sobre os aprendizados não são os mesmos em todos os setores do sistema educativo revela a necessidade de analisar separadamente, para cada contexto e setor do sistema educativo, quais os fatores que afetam o aprendizado. Este tipo de análise, em escala mais reduzida e de caráter mais contextualizado, deveria ter prioridade sobre as tentativas de construir um modelo explicativo com pretensões de universalidade, ou de formular medidas de política educativa de caráter geral para todo o sistema, derivadas diretamente dos “fatores associados”.

---

<sup>28</sup> Mais recentemente, o SAEB desenvolveu trabalho de pesquisa empregando técnicas de análise em múltiplos níveis que permitem tratamento mais apropriado da questão dos fatores associados.

<sup>29</sup> Ver, a este respeito, Ravela, P., Wolfe, R., Valverde, G. e Esquivel, J.M., 2000.

**Quadro 4**  
**EQUADOR. Conclusões do estudo sobre fatores associados**

*“Em geral, poder-se-iam enunciar as seguintes observações:*

- *As divergências são maiores que as coincidências entre os fatores de maior incidência no resultado acadêmico, o que dificulta a execução de ações comuns para fortalecer os elementos de maior impacto. Neste sentido, impõe-se levar em conta as peculiaridades de cada contexto.*
- *A aplicação de políticas educativas gerais ver-se-ia muito restringida pelo fato de que as coincidências dos fatores mais importantes nos contextos analisados que correspondem aos âmbitos da Gestão Institucional ou Escolar, Currículo e Prática Pedagógica são muito escassas.*
- *Deve-se destacar a presença de fatores que influem fortemente no resultado acadêmico dos estudantes e que correspondem ao âmbito familiar, especialmente a influência na Língua e Comunicação que têm os recursos de leitura no lar (aparece nos dois regimes de escolaridade e nos três tipos de estabelecimentos)e, no caso de Matemática, a educação formal do casal de pais.*

*De outro lado, cabe anotar que os fatores afetos ao âmbito da política educativa e que não ocupam os primeiros lugares porque sua condição atual não explica de forma importante os resultados do aprendizado estudantil deveriam merecer maior atenção e estudo para que se possa estabelecer, através de pesquisas pertinentes, quais práticas pedagógicas, ações de gestão escolar ou institucional e desenho de currículo são realmente “eficazes”, capazes de produzir impacto significativo no resultado acadêmico dos estudantes”.*

*Ministerio de Educación, Ecuador, 1999. Factores Asociados al Logro Académico. Resumen de difusión*

### V. 3. Três carências de caráter técnico comuns na região

Antes de finalizar esta revisão dos relatórios de resultados das avaliações nacionais de aprendizado, é imprescindível assinalar três carências ou deficiências técnicas verificadas na maioria dos relatórios e que são relevantes do ponto de vista da informação ao leitor especializado e da abertura da qualidade da informação gerada ao escrutínio público.

- Praticamente nenhum dos relatórios analisados informa as *taxas de resposta* aos diferentes instrumentos aplicados, vale dizer, que proporção dos alunos incluídos no desenho da amostra ou registrados nas escolas – no caso de operações censitárias – efetivamente participou da avaliação. Normalmente há informações sobre a quantidade de alunos avaliados, mas não sobre que proporção representam do total de alunos que deveriam ter sido avaliados. Esta informação é crucial, dado que em muitos países registram-se importantes níveis de absentismo nas provas ou de falta de resposta aos instrumentos complementares.
- Do mesmo modo, praticamente nenhum dos relatórios analisado inclui estimativa das *margens de erro de amostra*, que são básicos para analisar a significância da informação apresentada.
- Uma terceira deficiência é a falta de informações sobre os *procedimentos seguidos para a equiparação dos instrumentos de medição*, nos casos em que se aplicam avaliações sucessivas às mesmas séries e disciplinas. Este tema é de enorme transcendência, já que remete à questão de saber-se se as diferenças registradas entre anos refletem mudanças reais ou deficiências de nossos instrumentos.

**Estas carências em aspectos relevantes de caráter técnico fazem pensar na necessidade de trabalhar conjuntamente na região pela definição de um conjunto de critérios e requisitos técnicos que todos os relatórios deveriam atender.** Para isto pode ser útil tomar como ponto de partida os

diversos padrões de medição e avaliação produzidos nos Estados Unidos por instituições como a *American Educational Research Association – AERA* (Associação Americana de Pesquisa Educativa), ou o *National Center of Educational Statistics – NCES* (Centro Nacional de Estatísticas Educacionais) do governo federal.

### Dez Sugestões Para Pensar o Futuro

Como forma de encerrar esta revisão dos relatórios de resultados das avaliações de aprendizado, formularemos, na seqüência, o que consideramos as dez principais sugestões para melhorar nossos modos de informar os resultados, assim como para melhorar nosso conhecimento sobre até que ponto nossos relatórios efetivamente estão tendo algum impacto no sistema educativo.

1. Uma primeira pista para reflexão é, na realidade, uma advertência a respeito do cuidado que se deve ter antes de atribuir às avaliações nacionais algum tipo de consequência direta para as escolas, como, por exemplo, incentivos econômicos, publicação de classificação e outros. Neste sentido seria pertinente realizar algum tipo de pesquisa de tipo qualitativo sobre o impacto que uma estratégia desta natureza realmente tem nas escolas de um país como o Chile, que apostou fortemente nesta linha. Como indicamos no primeiro capítulo, este enfoque gera uma forte pressão sobre as escolas no sentido de dedicar muito tempo à preparação de seus alunos para responderem às provas, o que pode dar lugar a uma perigosa redução do currículo quando as provas aplicadas forem exclusivamente de múltipla escolha ou de resposta simplificada.

2. Se o objetivo principal é que os sistemas de avaliação dêem aos professores informações relevantes para que revisem e melhorem seu trabalho, então deveríamos gerar conhecimento específico em torno das seguintes perguntas: a) os professores estão

recebendo os relatórios e publicações derivadas das avaliações nacionais? b) em que proporção efetivamente os lêem? d) Até que ponto as informações que lhes damos são pertinentes e enriquecem sua compreensão dos processos de aprendizado dos alunos e os ajudam na análise de suas próprias práticas de ensino? E, f) que outras expectativas ou demandas têm ou fazem os professores às Unidades de Avaliação?

3. Mesmo que não se deseje seguir o caminho da publicação na imprensa dos resultados por escola, permanece o desafio de como informar às famílias sobre os resultados das avaliações. As perguntas que devem ser colocadas neste ponto são: que tipo de informações lhes devemos dar e de que modo devemos fazê-lo. Em princípio deveria tratar-se, mais que de informações numéricas – dados de resultados – de informações ilustrativas acerca do que é que esperam que seus filhos sejam capazes de fazer, até que ponto o estão conseguindo e o que podem fazer as próprias famílias para apoiar o trabalho da escola.

4. No mesmo sentido, seria útil saber sobre como outros atores relevantes – autoridades educativas, profissionais de unidades de currículo ou formação de professores, políticos, dirigentes sociais – compreendem e interpretam os resultados e que demandas de informação teriam a fazer aos sistemas de avaliação. Um modo de obter respostas para as perguntas formuladas neste parágrafo e nos dois que o precederam, seria realizar estudos de casos nacionais através de pesquisas e grupos de discussão com professores, famílias e outros atores relevantes, com o fim de indagar sobre suas percepções e demandas sobre os sistemas de avaliação de aprendizado.

5. Seria preciso fazer uma revisão sistemática de como a imprensa vem divulgando os resultados das avaliações nos países da região, o tipo de visão que se deriva para a opinião pública e o tipo de erros ou interpretações inadequadas mais comuns. Eventualmente, dever-se-ia pensar em instâncias formativas dirigidas a jornalistas dedicados a temas educativos (a Argentina

começou a fazê-lo recentemente, convidando um especialista em espanhol a apresentar um seminário sobre avaliação para jornalistas).

6. Para responder ao que os diferentes atores esperam dos sistemas de avaliação, provavelmente será necessário avançar mais em direção a formas de desenho de provas e de relatórios de resultados de caráter “referido a critérios”, vale dizer, que não se limitem a descrever médias de respostas corretas e a comparar tipos de escolas ou jurisdições em que o país está dividido, mas que estabeleçam um certo grau ou nível, e que informem até que ponto foi alcançado. Além disto, os relatórios numéricos deveriam ilustrar amplamente o que significam as pontuações em termos de conhecimentos e competências dos alunos. Estas ilustrações deveriam ir além da publicação de exemplos de questões individuais.

7. É preciso incorporar de alguma maneira a consideração da composição sociocultural dos alunos dos diferentes tipos de escolas ou setores do sistema educativo na análise e na apresentação dos resultados.

8. Seria preciso “agregar valor” aos relatórios, em termos de reflexão sobre os dados e sobre a relação entre eles e o que outras pesquisas revelam, entre outras coisas. Além disto, é preciso informar sistematicamente outras variáveis sociais e institucionais em si, de modo a enriquecer a compreensão de como de fato é o sistema educativo, e de sua heterogeneidade interna. Isto deveria ser feito, inclusive, antes de incorporá-las aos estudos dos “fatores associados”, e da maneira de medi-los. Por outro lado, é preciso avançar, e encomendar pesquisas que assumam adequadamente a complexidade metodológica do tema e moderar as expectativas sobre as contribuições que tais pesquisas poderiam dar para a formulação de políticas educativas, evitando-se cair em simplificações. É conveniente também ampliar o espectro das pesquisas sobre “fatores associados”, incluindo abordagens de caráter qualitativo.

9. Um desafio relevante que precisa ser assumido é a monitoração dos resultados de

nossos sistemas educativos ao longo do tempo. Este talvez seja um dos tipos de informações relevantes, tanto para o público em geral, como para as autoridades e tomadores de decisão. Esta tarefa tem fortes implicações de caráter técnico para garantia da equivalência dos instrumentos ao longo do tempo.

10. Finalmente, as unidades de avaliação da região teriam de trabalhar de forma

conjunta na definição de padrões ou parâmetros de qualidade técnica que as avaliações e os relatórios de resultados deveriam satisfazer. Seguramente este seria um passo enriquecedor para todos e permitiria melhorar a qualidade do serviço que nós, as Unidades de Avaliação, prestamos ao sistema educativo e à sociedade.

## BIBLIOGRAFIA

- BEATON, A., (1999. ). *International assessments: the United States TIMSS experience* In: F. Ottobre (ed.), *The role of measurement and evaluation in education policy*. Paris, UNESCO Publishing, Educational studies and documents N° 69.
- BENVENISTE, L., (2000). *La Evaluación del Rendimiento Académico y la Construcción de Consensos en Uruguay*. Lima: PREAL/GRADE.
- CARNOY, M., (2000). *School Choice? Or is it Privatization?*. In: *Educational Researcher*, Vol. 9, N° 7, pp. 15-20.
- LLACH, J.J., MONTOYA, S. e ROLDÁN, F. (2000). *Educación para Todos*. DISTAL, Buenos Aires.
- LINN, R. & GRONLUND, N. (2000). *Measurement and Assessment in Teaching*. Prentice-Hall, 8ª edición.
- LINN, R., (2000). *Assessments and Accountability*. In: *Educational Researcher*, Vol. 29, N° 2, pp. 4-16.
- MESSICK, S., (1999). *Key Issues*. In: F. Ottobre (ed.), *The role of measurement and evaluation in education policy*. Paris, UNESCO Publishing, Educational studies and documents N° 69.
- MINISTERIO DE EDUCACIÓN, República de Chile (2000). *Evaluación de Desempeño Establecimientos Educacionales Subvencionados. SNED 2000-2001*. Santiago.
- OTTOBRE, F., (ed.) (1999), *The role of measurement and evaluation in education policy*. Paris, UNESCO Publishing, Educational studies and documents N° 69.
- RAVELA, P. (editor), WOLFE, R., VALVERDE, G. e ESQUIVEL, J.M., (2000). *Los Próximos Pasos: ¿hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?* PREAL/GRADE, Lima.
- SECRETARÍA DE EDUCACIÓN PÚBLICA DE MÉXICO (SEP) - SINDICATO NACIONAL DE TRABAJADORES DE LA EDUCACIÓN (SNTE) (1994). *Carrera Magisterial. Folleto Informativo. Primera Vertiente*. SEP, México D.F.
- TUIJNMAN, A. e POSTLETHWAITE, N. (editores), (1994); *Monitoring the Standards of Education*. Pergamon, Oxford.
- UNESCO/OREALC, (1999). *Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado*. UNESCO/OREALC, Santiago.