

PROPÓSITOS DA AVALIAÇÃO EDUCACIONAL: uma análise das alternativas

PHILIP R. FLETCHER¹

1.0 Escola pública e ensino de boa qualidade: fundamentos

Um sentimento de angústia paira por cima do ensino público deste país. É um sentimento bastante generalizado de que se gasta mal em educação, priorizando projetos assistencialistas que pouco ou nada têm a ver com os propósitos educacionais. Após o abandono da escola pública pela classe média e a efetiva duplicação da rede pública com recursos próprios, a disposição do público em bancar maiores aventuras nessa área já está quase no limite. Antes de se pensar em mais recursos para a área do ensino, o sentimento é que se deva gastar racionalmente os recursos já existentes. Antes de mais nada, precisa-se entender que a função principal da escola é ensinar e que o resultado a ser esperado, avaliado e cobrado deve ser os resultados da aprendizagem do aluno. O público espera que as Secretarias de Estado da Educação forneçam evidências convincentes de que as escolas públicas funcionam plenamente, cumpram os preceitos da Constituição e ofereçam um ensino fundamental de boa qualidade; cobra das Câmaras Municipais, Legislaturas Estaduais e congêneres federais demonstração de que os fundos comprometidos com o ensino público promovem a aprendizagem dos alunos com o mesmo desempenho das melhores escolas particulares; pede um sistema de prestação de contas e incentivos que reconheça as reais diferenças nas funções, competências e desempenhos dos professores, premie as escolas com disposição para o progresso e, ainda, um

1 Da Fundação Carlos Chagas.

sistema de "inadimplência escolar" para aquelas que se demonstrem incapazes de prestar esses serviços em benefício dos alunos.

Tudo isso preconiza um sistema extensivo de aferição e um corpo administrativo que aproveite plenamente de seus resultados. Pressupõe-se a existência de critérios, instrumentação, medidas, técnicas de análise, mecanismos administrativos e procedimentos de reciclagem e de recuperação. No entanto, o Ministério da Educação, as Secretarias de Estado da Educação, Superintendências, Núcleos ou Delegacias Regionais de Ensino e as escolas públicas individuais têm parca experiência com o planejamento e implementação de verdadeiros sistemas de monitoramento. É necessário reconhecer que a prática da pesquisa em campo, dominada pelas universidades e certas instituições envolvidas com levantamentos sociais, fornece experiência relevante para apenas alguns dos componentes de um sistema dessa natureza. *Há uma exigüidade muito grande de estudos na área das políticas públicas que esclareçam a relação que existe entre os requisitos de informação de um sistema de aferição e o uso dessa informação no processo de execução das políticas públicas por aqueles que são responsáveis pela distribuição dos recursos ou a eventual aplicação das sanções.* A escolha dos instrumentos de avaliação corre em função desses objetivos. Uma vez que se esclareçam os objetivos, há uma ampla gama de inovações técnicas que deveriam ser consideradas na elaboração de um verdadeiro programa de avaliação educacional.

2.0 Pesquisa e avaliação: as bases escolar e domiciliar

Cabe lembrar que pesquisa não é avaliação. A pesquisa parte de uma base de amostragem para tirar *inferências* sobre um universo desconhecido. A avaliação, ao preconizar e antever ações *administrativas*, assume compromissos que vão muito além do relatório apresentado como produto final de uma pesquisa. A pesquisa convencional é tópica e intermitente, despreocupada com a responsabilidade imediata de solucionar as questões por ela investigadas. A avaliação, como atividade permanente e recorrente, pressupõe a existência de um relacionamento mais formal e duradouro com seus clientes institucionais e pessoais, envolve-se numa relação recíproca comprometida com a busca de eventuais soluções que seriam mutuamente satisfatórias e procura a atribuição clara das responsabilidades individuais em cada instância. Não há notícia de um sistema de avaliação digno do nome que opere exclusivamente na base da amostragem. Não há como justificar com argumentos da avaliação a participação de alguns poucos casos ao custo de ignorar a grande maioria, como ocorre no caso das pesquisas realizadas por amostragem. Ou alguém participa de um sistema de avaliação ou se exclui. Não é permitido que um componente entre no sistema de avaliação em determinado ano e sai no ano seguinte. O monitoramento envolve um compromisso a longo prazo, senão permanente.

No entanto, processos de longo prazo e pouca emoção, como a reestruturação das redes públicas de educação e de saúde, requerem maior

atenção do público para se tornarem sustentáveis. É preciso, porém, que temas dessa natureza, que não costumam liberar a adrenalina do público, sejam tratados de maneira a torná-los interessantes e claros. Isso exige, acima de tudo, inteligência. Caso o propósito imediato do governo seja o de agitar o país, sacudir as pessoas de sua habitual sonolência e colocar a educação num lugar de destaque no centro da agenda política nacional, a pesquisa ainda representa uma opção bastante atraente nesse mesmo sentido, porque promete o maior impacto sobre a opinião pública ao menor preço. A seu respeito, vale a pena considerar, mesmo que brevemente, as vantagens relativas das pesquisas educacionais de base escolar e de base domiciliar, quando são trabalhadas com inteligência, competência e imaginação.

2.1 Pesquisas de base escolar

Uma vantagem da pesquisa de *base escolar* é que gera resultados agregados que caracterizam as escolas que entram na amostra. No entanto, essas medidas do desempenho escolar devem ser relacionadas com as outras características da escola na procura das causas do bom desempenho e não para responsabilizar unidades administrativas específicas. As desvantagens dessa modalidade de pesquisa incluem a interrupção das atividades escolares e o fato de que nem todo mundo facilita o ingresso no recinto da escola para fins de pesquisa ou de avaliação. No Brasil, as escolas privadas são particularmente resistentes a propósitos desse tipo. Nem todo mundo em idade escolar frequenta escola. Por definição, a população que se encontra fora da escola é excluída das pesquisas de base escolar. Isto cria certos problemas técnicos na hora da análise dos resultados. A diferença entre o desempenho médio na 5ª e 4ª séries não representa, necessariamente, o crescimento cognitivo no intervalo da série por causa do abandono da escola após a 4ª série, o que abrange, evidentemente, alunos que fogem do âmbito da pesquisa e, portanto, não podem ser recuperados.

Adolescentes jovens e crianças freqüentemente não conseguem responder a perguntas sobre as condições sócio-econômicas do domicílio. Muitos não conseguem sequer identificar a profissão do pai ou responsável pelo próprio domicílio em que moram. O pesquisador pode cobrar uma resposta do aluno nesse sentido, mas essas respostas são notórias pela sua falta de fidedignidade. É evidente, então, que quaisquer temas que envolvam considerações de nível sócio-econômico ou de equidade na distribuição das oportunidades de ensino fogem do alcance das pesquisas de base escolar. A não ser que um inquérito desse tipo seja complementado por entrevistas no domicílio de cada aluno, é praticamente impossível aprofundar a apuração das causas endógenas do bom desempenho acadêmico devido à interferência de disparidades sócio-econômicas exógenas à escola. Ao tocar na relação da escola com a sociedade maior na qual ela se encerra, deve-se reconhecer que as pesquisas de base escolar não conseguem levantar qualquer informação a respeito da população

economicamente ativa após a escolarização. As investigações sobre a suficiência da escolarização recebida, quando aplicadas nos diversos setores da economia, e, especialmente, em relação à questão das qualificações e habilidades exigidas pelas novas tecnologias, fogem do alcance das pesquisas de base escolar.

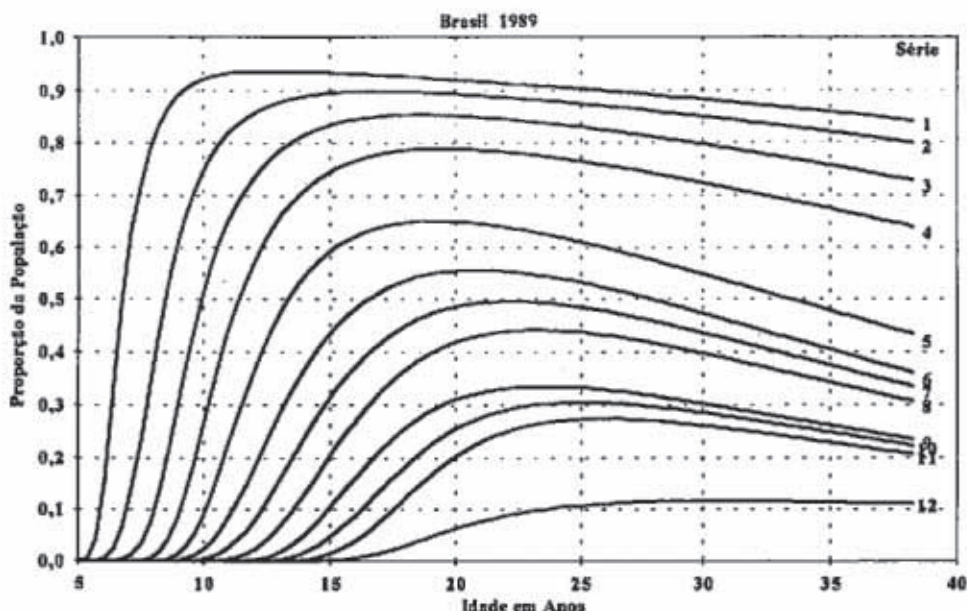
2.2 Pesquisas de base domiciliar

Se o propósito do governo é saber como anda o ensino do país para alimentar considerações sobre políticas públicas nesse setor, a pesquisa de base domiciliar oferece uma série de vantagens sobre a alternativa de pesquisar escolas. A desvantagem é que ignora as características internas das escolas, especialmente sua organização e métodos, a não ser que seja complementada por entrevistas nas escolas. No entanto, a pesquisa de base domiciliar não interrompe a rotina diurna da escola e não depende da permissão concedida pelos donos de escola, diretores e professores para levantar informações sobre o perfil cognitivo da população. Basta obter essas informações diretamente no domicílio, através de entrevistas aplicadas nas faixas etárias de interesse. A presença no domicílio facilita a coleta de informações pormenorizadas e fidedignas sobre os rendimentos do domicílio e as reais condições de vida dos habitantes. A base domiciliar assegura que se encontrem juntos alunos de escolas públicas e particulares (embora as escolas nem sempre possam ser identificadas nominalmente), a população em idade escolar dentro e fora da escola e, ainda, a população economicamente ativa nas faixas etárias investigadas, permitindo que os perfis cognitivos desses grupos sejam levantados e analisados. Quando toda a base demográfica está presente na análise, pode-se reconstituir os processos de transição das séries e, ainda, da transição *de um estado cognitivo para outro*, nos diversos segmentos da população investigada.

O Gráfico 1 mostra as taxas de participação no ensino de 1º e 2º graus do Brasil em 1989 (onde as séries do 2º grau são enumeradas de 9 a 11 e o ingresso no ensino superior é representado por 12), ou seja, a proporção da população que ingressa em cada série de ensino. Com as técnicas do modelo Profluxo aplicadas aos dados de base domiciliar, pode-se reconstituir todos os detalhes do fluxo escolar, inclusive as taxas de repetência, promoção e evasão que compõem a matriz de transição de série. Transportado para o âmbito cognitivo, percebe-se que o Gráfico 1 serve como uma analogia particularmente sugestiva. Se as curvas representassem os *níveis de desempenho cognitivo*, em vez das séries de ensino formal, poder-se-ia retratar o percurso ao longo dos diversos níveis de aprendizagem, desde a infância até a vida adulta.

Na *subida* das curvas de contorno cognitivas, seria possível calcular a idade média do domínio de um determinado nível de competência; o que levaria a considerações sobre o *timing* de uma seqüência de eventos cognitivos. No *ponto máximo* dessas curvas, poder-se-ia determinar o desempenho atual do

GRAFICO 1
Taxas de Participação, segundo Série e Idade



sistema escolar em termos de sua cobertura da população, ou seja, a proporção da população que atinge certo nível cognitivo, permitindo reflexões sobre a suficiência da escolarização recebida. Além do máximo dessas curvas, retratar-se-ia o desempenho do sistema escolar em termos dos resultados cognitivos obtidos em passado recente, o que forneceria uma perspectiva longitudinal do sistema educacional, útil para projeções do desempenho cognitivo no futuro. Ao sobrepor os resultados da avaliação cognitiva em cima das taxas de participação em cada série, seria possível determinar o que se aprende, em média, em cada série de ensino. Seria, então, possível examinar como a aprendizagem dos alunos e os critérios de promoção condicionam a transição de série e a evasão, estruturando as oportunidades do ensino e, em última análise, a distribuição de escolarização na população. Estas são apenas algumas das potencialidades a serem exploradas com a pesquisa de base domiciliar.

3.0 Avaliação dos sistemas educacionais: componentes

Caso o propósito do governo seja o de estabelecer um relacionamento mais estreito com a clientela das escolas e de assumir um compromisso de parceria na busca de eventuais soluções, a avaliação tem, evidentemente, um alcance muito maior que a pesquisa, embora a custos também muito mais elevados. A avaliação, como sistema de atividades, envolve três componentes estreitamente relacionados: (a) um levantamento periódico de informações; (b) a análise e

(c) o aproveitamento dessas informações para orientar eventuais ações administrativas. Ou seja, *um sistema de monitoramento requer um levantamento recorrente de informações a intervalos regulares e previsíveis com o propósito de produzir um juízo de valor sobre a condição, estado, direção ou taxa de crescimento de um sistema ou um de seus componentes no contexto das responsabilidades assumidas por uma cadeia administrativa de ação.*

Quando critérios, informações e atribuições são claros, as pessoas por eles enquadradas conseguem se liberar das arbitrariedades administrativas e interferências indevidas. Nesse contexto, a avaliação representa um reforço positivo que afasta insinuações e dúvidas, potenciais fontes de insegurança, angústia e tensão. Dessa forma, a avaliação não seria jamais inquisitorial. Pelo contrário, pode permitir uma maior descentralização das atividades e ampliar a margem de autonomia e liberdade de ação das pessoas. O sentimento de pânico que costuma acompanhar uma investida administrativa sumária cede lugar a um relacionamento governado por procedimentos periódicos previsíveis, no interesse da promoção e progressiva melhoria das pessoas e das instituições. Quando a avaliação revela desmandos e deficiências, na grande maioria dos casos, as ações administrativas mais rigorosas permanecem apenas iminentes. A iminência da retribuição é suficiente como controle administrativo na grande maioria das vezes, porque, normalmente, as pessoas conseguem tirar as devidas conclusões *quando* informadas sobre seu desempenho. Apenas muito eventualmente, e quando ações mais enérgicas se tornam inadiáveis, as investidas administrativas seriam necessariamente comprometidas com a busca de soluções mutuamente satisfatórias e implementadas com a devida transparência, exatamente *por causa* do rigor e clareza dos procedimentos da avaliação.

4.0 Síntese de algumas experiências

Existem diversas maneiras de classificar os sistemas de monitoramento educacional. Apresento um esquema apenas para demonstrar que a avaliação exige *compromisso*, se bem que ainda exista dúvida quanto a isso. Descrevo, a seguir, três protótipos de sistemas de monitoramento, conforme existiam nos Estados norte-americanos de Nova Jérsei, Minnesota e Califórnia, no início da década de 80 (Tabela I). Cada protótipo representa um tipo idealizado de sistema de avaliação que corresponde à principal iniciativa que dominava cada um desses Estados naquele momento, sem necessariamente ser o único sistema por eles endossado ou aplicado. Meu propósito é de contribuir para a discussão das alternativas de monitoramento ora em consideração no Brasil e não o de oferecer um retrato totalmente fiel ao registro histórico.

Tabela I

Uma Tipologia de Sistemas de Monitoramento

Distinção	Tipo		
	Conformidade	Diagnose	Rendimento
Referencial	Organização Burocrática	Psicologia Clínica	Concorrência Econômica
Objetivo	Padronizar o Sistema	Remediar Deficiências de Aprendizagem	Incentivar o Desempenho Acadêmico
Superintendência	Insumos	Procedimentos	Resultados
Exemplo	Nova Jérsei"	"Minnesota"	"Califórnia"

4.1 Avaliar os insumos: padronizar o sistema

O tipo mais antigo dos sistemas de monitoramento verifica a conformidade da escola com os padrões técnicos especificados pelas agências públicas. Ao certificar a conformidade, tenta-se assegurar que a escola alcance e mantenha um padrão predeterminado. Essa modalidade de monitoramento se distingue das outras modalidades ao focalizar, quase que exclusivamente, os insumos da escola. Verifica-se, por exemplo, que as credenciais dos professores foram devidamente examinadas e arquivadas; que os metros quadrados da biblioteca estejam de acordo com as especificações para escolas desse tamanho; que os extintores de incêndio foram devidamente recarregados e inspecionados. Esse padrão de conformidade é definido por uma série de leis, regulamentos e códigos administrativos endossados pelas diversas instâncias de governo. De acordo com esses princípios, ou a escola está em conformidade com os requisitos ou não está. Às vezes, isto é verificado mediante uma simples auto-declaração assinada pelo diretor da escola. Em outros casos, as declarações recebem o complemento de uma inspeção no local.

Quando a escola for notificada de que não está em conformidade, o governo dispõe de três alternativas de ação: (a) a escola pode receber um atestado acusando a falta de conformidade; (b) a escola recebe um prazo limitado para corrigir as deficiências; ou (c) um certificado é emitido livrando a escola da responsabilidade de entrar em conformidade com os regulamentos. A primeira dessas ações geraria apreensões públicas e dificilmente seria implementada na prática. É mais provável que um prazo seja definido para corrigir o problema. Mas, geralmente, com o acúmulo das informações que atestam a inconformidade, a escola sofre aumentos progressivos na supervisão pelas agências educacionais públicas. Outra consequência pode ser a eventual retirada dos recursos estaduais da escola ou a perda da licença de operação concedida pelo Estado.

O Estado de Nova Jérsei operava um sistema desse tipo na década de 80. Nesse Estado, a escola recebe uma notificação prévia especificando a data e os requisitos da avaliação. Os regulamentos exigem da escola uma relação de objetivos educacionais escritos por extenso – um dos quais teria que promover o desempenho acadêmico do aluno – e um plano organizacional quinquenal. Outros elementos incluem o registro diário da frequência dos alunos; a aplicação de provas que avaliam habilidades básicas; currículo e instrução compreensivos; contabilidade escolar; relações escola-comunidade; certificação da planta física da escola e um plano de carreira para professores e supervisores. Na base da avaliação, a escola recebe uma certificação de Nível I, II ou III, um esquema de pontuação graduada que acelera a frequência das avaliações e aumenta a intervenção do Estado. A partir daí, o governo obriga a escola a rever a avaliação externa baseada nos documentos recebidos e exige da escola um plano de ações para implementar as correções e exigências da equipe de revista. Nesse ponto, a escola pode entrar na justiça com recursos para rever suas obrigações à luz das recomendações. Nos casos mais difíceis, o Estado pode intervir e administrar diretamente a escola por um período de até cinco anos.

4.2 Avaliar os processos: remediar deficiências de aprendizagem

Na segunda modalidade de monitoramento, a dinâmica do processo de ensino-aprendizagem torna-se o foco das atenções, onde o objetivo é a progressiva melhoria do aluno mediante diagnoses periódicas vinculadas a procedimentos de recuperação. O instrumento diagnóstico de preferência é a prova referenciada a critérios, bem articulada com o conteúdo curricular da escola. A diagnose segue uma seqüência específica. O aluno faz uma prova para avaliar seu conhecimento antes do início do ensino. O professor orienta o ensino para os objetivos curriculares ainda não consolidados. No final da unidade curricular, o aluno realiza outra prova de igual teor. O professor desenvolve um plano de recuperação baseado nos resultados da segunda prova, o que inclui aulas adicionais, exercícios e a aplicação de outras provas ainda. Em princípio, no final desse processo, todos os alunos devem dominar os objetivos curriculares.

Um exemplo dessa modalidade de monitoramento opera no Estado de Minnesota, consistentemente um dos primeiros colocados nas provas da *National Assessment for Educational Progress* (NAEP), conforme mostra a Tabela II. O sucesso relativo do sistema educacional de Minnesota alimenta uma reticência em ceder às pressões de reforma que tomou conta dos outros Estados norte-americanos na década de 80. Autoridades locais preferem confiar nas potencialidades e experiência das escolas locais e resistem à tentação de comparar essas escolas em público. Ao suplementar as iniciativas de avaliação locais, o Estado aplica provas em todo seu território em quinze áreas temáticas diferentes, custeadas pelo governo estadual. Mas os resultados dessas provas são tratados com confidencialidade entre professores, alunos e pais, como seria no caso de um relacionamento clínico envolvendo médicos e pacientes.

Tabela II
Nível de Desempenho em Matemática^e

População	Escore		Porcentagem que Alcança ou Supera o Nível		
	Médio	Desvio	"Básico" 256 ^f	"Proficiente" 294 ^g	"Adiantado" 331 ^h
China ^{a,d}	291	28	87,7	45,4	7,8
Taiwan ^{a,c}	285	38	78,0	40,7	10,4
Coréia ^{a,c}	283	32	80,6	36,9	6,4
MINNESOTA, EUA ^b	282	32	79,0	37,0	6,0
Suíça ^{a,c}	282	26	83,9	33,1	3,4
União Soviética ^{a,c}	279	28	79,7	30,0	3,1
Hungria ^{a,c}	277	29	76,6	28,6	3,3
França ^{a,c}	273	28	72,6	23,2	2,0
Emília-Romanha, Itália ^{a,c}	272	28	72,2	22,1	1,7
Israel ^{a,c}	272	27	72,5	21,6	1,5
NOVA JÉRSEI, EUA ^b	272	35	67,0	28,0	4,0
Canadá ^{a,c}	270	27	70,3	19,4	1,4
Escócia ^{a,c}	270	28	69,0	19,5	1,4
Inglaterra ^{a,d}	270	29	68,1	20,6	1,9
Irlanda ^{a,c}	269	28	67,7	19,2	1,5
Eslovênia ^{a,c}	266	27	64,5	15,9	1,0
Espanha ^{a,c}	262	26	59,6	11,7	0,5
Estados Unidos ^{a,c}	262	29	58,1	14,2	1,3
CALIFÓRNIA, EUA ^b	261	37	55,0	20,0	3,0
Portugal ^{a,d}	258	26	53,2	9,4	0,4
Jordânia ^{a,c}	248	27	37,8	5,6	0,2
São Paulo, Brasil ^{a,d}	243	27	30,7	4,1	0,2
Fortaleza, Brasil ^{a,d}	240	26	25,8	2,9	0,0
Maputo e Beira, Moçambique ^{a,d}	239	20	18,8	0,8	0,0

^a Escolas públicas e particulares, alunos de 13 anos. 1991 International Assessment of Education Progress (IAEP). P.J. Pashley & G.W. Philips, *Toward World-Class Standards: A Research Study Linking International and National Assessments* (Princeton, N.J.: Educational Testing Service, 1993, Tables 4 & 5, p. 25-6.

^b Escolas públicas, alunos da 8ª série. 1992 National Assessment of Education Progress (NAEP). I.V.S. Mullis, J.A. Dossey, E.H. Owen & G.W. Philips, *NAEP 1992 Mathematics Report Card for the Nation and the States* (Washington, D.C.: U.S. Department of Education, 1993), Table 4, p. 10.

^c Populações compreensivas.

^d Populações com exclusões ou participação baixa.

^e As cinco áreas de conteúdo examinadas incluem (1) números e operações, (2) medidas, (3) geometria, (4) análise de dados, estatística e probabilidade, e (5) álgebra e funções. Os três níveis de desempenho foram identificados pelo National Assessment Governing Board (NAGB), que supervisiona o NAEP. Os resultados comparativos internacionais foram obtidos com técnicas de regressão aplicadas aos dados do IAEP e NAEP (Pashley & Philips, 1993). Escores médios e desvios padrão estimados pelo autor.

^f No nível de desempenho "básico", os alunos demonstram compreensão conceitual e o domínio de procedimentos nas cinco áreas de conteúdo.

^g No nível de desempenho "proficiente", os alunos aplicam conceitos e procedimentos à resolução de problemas complexos nas cinco áreas de conteúdo.

^h No nível de desempenho "adiantado", os alunos vão além do reconhecimento, identificação e aplicação dos princípios matemáticos e conseguem generalizar e sintetizar conceitos e procedimentos nas cinco áreas de conteúdo.

Professores e supervisores locais participam da escolha dos instrumentos, definição dos critérios, interpretação dos resultados, comparações do desempenho de grupos e na identificação e uso dos recursos para melhorar o ensino. Mais de 70% dos professores declararam que continuariam usando as provas, mesmo que não fossem compulsórias; mais de 80% afirmaram que os resultados das provas levaram a alterações nas políticas educacionais locais; e mais de 85% disseram que os resultados do programa de avaliação foram consultados na reformulação anual do currículo. Os níveis de aceitação, participação e comprometimento são altos, característicos de um corpo técnico bem treinado, bem motivado e bem remunerado.

4.3 Avaliar o rendimento: incentivar o desempenho acadêmico

O monitoramento do rendimento representa ainda uma terceira opção. Embora exista há muitos anos, essa modalidade de avaliação recebeu um novo ímpeto na década de 80, como parte de um movimento para maior transparência na administração das escolas públicas, mediante a periódica prestação pública de contas (*accountability movement*). Pressupõe-se que a demonstração do rendimento das escolas bem sucedidas sirva para estimular a melhoria do desempenho nas outras escolas. Tipicamente, esse tipo de avaliação prioriza o uso das provas cognitivas, em que os resultados são agregados por escola ou distrito escolar antes de torná-los públicos. Na década passada, o monitoramento do desempenho foi implementado num grande número de áreas urbanas; no entanto, o Estado da Califórnia ganhou notoriedade por ter desenvolvido um sistema de alcance e uniformidade estaduais.

O sistema de avaliação do ensino da Califórnia resultou de uma batalha política travada entre o governador e o superintendente estadual de ensino. O governador contestou em público uma proposta para aumentar as verbas do ensino na ausência de qualquer evidência convincente de que se conseguiria melhorar o desempenho com esse acréscimo de recursos. O superintendente, para angariar esses recursos, respondeu às preocupações do governador e estabeleceu um sistema anual de avaliação do ensino. A finalidade do *Relatório de Desempenho* anual é propor objetivos e patamares de desempenho para orientar a melhoria progressiva dos professores, alunos e escolas.

Três antecedentes críticos contribuíram para o êxito do sistema de prestação de contas implementado na década de 80. O primeiro, envolveu a consolidação de um sistema que padronizou a coleta de dados e a tornou rotina, evitando a duplicação e a casualidade com que essas informações foram tratadas anteriormente. Uma vez por ano, no penúltimo mês do ano letivo, a secretaria estadual de ensino obriga professores e diretores de escolas a reservar um dia inteiro de serviço exclusivamente para fornecer informações sobre escolas, pessoal, matrículas e oportunidades letivas em formulários fornecidos pelo Estado.

Da mesma forma, o desempenho dos alunos foi medido anteriormente por uma variedade de provas aplicadas localmente com uma cobertura variável da população nas diferentes séries de ensino. Desse jeito, as medidas de rendimento não permitiam comparações entre escolas. Assim, a consolidação de um programa estadual de provas representou a primeira tentativa de avaliar o desempenho cognitivo dos alunos com base em um conjunto uniforme de requisitos, instrumentos e procedimentos. Esse sistema fornece dados sobre o desempenho dos alunos no ensino público da 3ª, 6ª, 8ª e 12ª séries. O terceiro elemento construtivo veio das instituições do ensino superior público do Estado, que publicam relatórios sobre o desempenho de seus estudantes universitários. Essa informação inclui as notas médias dos estudantes no primeiro ano do ensino universitário, junto com os resultados em provas de aptidão numérica e verbal.

Juntos, esses três sistemas fornecem a base para um relatório anual sobre o desempenho das escolas no Estado da Califórnia. Esse relatório contém cinco capítulos – quatro sobre a qualidade do ensino e um quinto com informações sobre o contexto social das escolas e de seus alunos. Os indicadores de qualidade incluem: (a) estatísticas sobre a matrícula segundo a matéria e sobre os requisitos de promoção; (b) distribuição dos resultados cognitivos para a 12ª série em matemática e compreensão de leitura com a apresentação dos 1º, 2º e 3º quartis, além da média da escola; (c) o desempenho no vestibular dos alunos aceitos para ingresso no ensino superior e resultados das provas de desempenho nas cinco áreas examinadas com maior frequência; e (d) informações sobre frequências e desistências. As informações contextuais contêm referências ao domínio do inglês pelos alunos e à composição étnica dos alunos da escola, entre outras informações.

O Estado também fornece assistência técnica para o levantamento de informações locais. As recomendações solicitam informações sobre (a) as opções curriculares da escola; (b) percepção do ambiente da escola; (c) quantidade e qualidade de trabalhos escritos pelos alunos; (d) quantidade e qualidade dos deveres de casa; (e) número e tipo de livros lidos pelos alunos; (f) apoio comunitário recebido pela escola; (g) honras e prêmios recebidos pelos alunos; (h) natureza e qualidade do apoio oferecido a alunos que demonstram necessidades especiais no ensino; e (i) participação dos alunos em atividades extracurriculares.

Para fins de comparação, as escolas são ordenadas de acordo com um escore que representa o desempenho cognitivo dos alunos. Procedem-se, então, ao agrupamento das escolas numa "faixa flutuante" com 80 escolas acima e abaixo da escola, permitindo, ainda, que a escola migre de faixa de um ano para outro. Para ajudar as escolas na seleção de seus objetivos acadêmicos, o Estado calcula escores padronizados dos indicadores para as escolas em cada faixa. O relatório identifica para cada indicador os escores ou valores que a escola deveria obter para alcançar ou superar o 75º percentil de seu grupo de referência.

4.4 Diversas modalidades de avaliação: limitações e controvérsias

Conformidade, diagnose e desempenho no monitoramento do ensino se referem a diferentes partes do sistema escolar e têm referenciais teóricos e objetivos diferentes. O monitoramento da conformidade tem sua origem na literatura sobre regulamentação e controle burocrático. Seu objetivo é o de estabelecer e manter padrões endossados pelo poder público, examinando os insumos do sistema educacional. Isto pressupõe que existam insumos suficientes para uma educação de qualidade e que esses requisitos possam ser padronizados.

Como sistema de avaliação, o monitoramento da conformidade é uma opção que minimiza custos e as necessidades de recursos humanos. Onde o alcance da autoridade é limitado, quando as secretarias estaduais de ensino não dispõem de quadros altamente qualificados, e onde o controle da escola é francamente hostil, o monitoramento da conformidade ainda pode representar a melhor opção possível nas circunstâncias. Sua força é chamar a atenção para os casos mais agudos, bem abaixo do padrão, mas dificilmente essa modalidade de avaliação dá conta das diferenças que surgem entre alunos e escolas.

A aplicação de padrões de conformidade uniformes em todo um Estado, sem considerar o contexto social no qual a escola se insere, pode resultar numa distribuição inadequada dos recursos. Também promove a ilusão de que todas as escolas encontradas em conformidade forneçam uma educação de qualidade igual. Exemplos incluem o de uma escola situada no outro lado da rua em que existe uma biblioteca pública e que, assim, não precisa manter uma biblioteca própria. Os recursos da biblioteca seriam melhor aplicados numa outra escola. Um exemplo de conformidade meramente formal seria a escola que tem um número de livros na biblioteca suficiente para a matrícula, mas os livros são todos desatualizados e irrelevantes em termos curriculares. Na prática, as variáveis que representam insumos quase nunca se correlacionam com o desempenho de alunos ou de escolas nas pesquisas educacionais. Pode ser que o monitoramento da conformidade seja adequado para satisfazer necessidades nas áreas de saúde e de segurança, mas parece menos eficaz no papel de promover o desempenho acadêmico.

Já o monitoramento dos processos de aprendizagem tem sua inspiração na psicologia clínica. Seu objetivo é fornecer uma diagnose adequada das deficiências de aprendizagem do aluno e de assegurar que receba adequadas oportunidades de atenção e de recuperação. O foco da superintendência do monitoramento é o processo de ensino-aprendizagem, um processo a ser aprimorado através da elaboração dos planos de ensino e do controle do ritmo da apresentação do currículo. A vantagem dessa modalidade é sua estreita relação com o currículo. Os alunos são examinados nas matérias que recebem ênfase no currículo e os processos de recuperação dão ênfase aos processos cognitivos identificados pela escola como essenciais para o desenvolvimento do aluno.

No entanto, um programa de diagnose e de recuperação requer um alto grau de entrosamento com o currículo da escola para alcançar seus objetivos. É difícil imaginar como esse tipo de sistema poderia ser usado em grande escala sem um currículo único endossado pelo poder público. O problema maior não seria exatamente a diagnose, mas sua integração com os processos de recuperação. Na prática, é difícil atender a demanda por recursos e professores adicionais nas escolas com grande contingente de alunos que precisam de ajuda.

O monitoramento diagnóstico segue uma orientação clínica que focaliza o aluno individualmente. O modelo tende a ignorar contingências exógenas à escola que podem esmagar a dinâmica do ensino em sala de aula. Os modelos clínicos ignoram fontes importantes da variação nos resultados do ensino ao desconsiderar a origem sócio-econômica do aluno, a liderança administrativa de elementos dentro da escola e, ainda, a própria mobilidade do aluno. Escolas com altos índices de transferência de alunos não conseguem utilizar plenamente esse método, devido às interrupções no ciclo de ensino, aprendizagem, provas e recuperação.

O monitoramento diagnóstico pressupõe uma orientação clínica que não se coaduna com o monitoramento do rendimento. Seria um desatino usar uma diagnose clínica para promover uma concorrência pública. O modelo diagnóstico já parte do pressuposto de que os alunos demonstrem habilidades e interesses diferentes. Embora se espere que os alunos alcancem os objetivos curriculares previstos, não há nenhuma expectativa de que isso aconteça de forma competitiva. A partir daí, fica evidente que nem todos os alunos aprendem tudo no mesmo ritmo. No entanto, quando os professores são obrigados a cobrir determinada matéria num prazo estipulado e, ao mesmo tempo, a fornecer amplas oportunidades de recuperação, eles se colocam no meio de conflitos curriculares que são difíceis de resolver sem o acréscimo de pessoal especializado e recursos adicionais para as atividades de recuperação.

O monitoramento do rendimento tenta aplicar os princípios da teoria microeconômica a organizações do setor público. O objetivo é promover o desempenho acadêmico através da concorrência entre unidades do sistema educacional. Seu foco privilegia os resultados cognitivos alcançados pela escola. Sistemas compreensivos de monitoramento do rendimento como o da Califórnia ainda são novidade no campo educacional. Percebe-se, no entanto, que existe uma nítida reticência em identificar exatamente *quais* incentivos deveriam ser associados às medidas de rendimento. Existem tentativas para associar o rendimento dos alunos nas provas à promoção dos professores no plano de carreira. Há ainda propostas de relacionar a avaliação dos professores ao rendimento dos alunos em sala de aula. No entanto, em praticamente todos esses casos, as tentativas não passam de experiências. Na maioria dos casos, espera-se usar a publicidade negativa associada a um desempenho deficiente para estimular administradores e professores a dedicarem maiores esforços à melhoria do desempenho dos alunos. Fica evidente, no entanto, que uma das

grandes alternativas na área do monitoramento do rendimento envolve a escolha dos tipos de sanções econômicas, políticas ou regulamentares a serem usados como incentivos.

Há diversos problemas com essa modalidade de avaliação. Ela depende de uma enorme melhoria na capacidade de processamento de informações que seria difícil de alcançar e manter, mesmo com a revolução que ocorre na informática, devido ao alto grau de entrosamento administrativo necessário. Na medida em que Estados e escolas não disponham do fôlego financeiro e competência técnica necessários para gerenciar um sistema de monitoramento do desempenho com os novos recursos da informática, este seria um sistema a ser evitado. Uma segunda desvantagem pode ser o eventual estreitamento do currículo por escolas que tentam se adaptar aos sistemas de monitoramento do rendimento. Ao enfrentar as limitações de recursos e o aumento das pressões políticas externas, pode ser que as escolas eliminem todo o ensino que não se relacione diretamente com a educação básica, incluindo cursos sobre arte e música nas escolas de 1º grau e as oportunidades mais avançadas em matemática e ciências no ensino de 2º grau.

Nem todo mundo tem condições de responder aos sinais do mercado preconizado pelo sistema de monitoramento. Exigências de residência no local reduzem as opções de escolha. Mesmo que os requisitos de residência fossem abandonados, crianças pequenas não teriam a liberdade dos pais para atravessar a cidade na procura do ensino de alta qualidade. Nas circunstâncias de monopólio ou de oligopólio, os pressupostos teóricos embasados nos princípios de um mercado livre não se aplicam. Precisa-se lembrar, também, que recursos financeiros, talento de liderança administrativa, aptidão para o ensino e condições físicas não são compartilhados igualmente entre todas as escolas. É, portanto, duvidoso pressupor que as unidades deficientes tenham iguais oportunidades para responder aos incentivos e sanções de um sistema de monitoramento do rendimento. Pode ser, inclusive, que uma das conseqüências indevidas desse tipo de monitoramento seja a de exacerbar as diferenças cognitivas das comunidades ricas e pobres, urbanas e rurais, reforçando as disparidades sociais existentes.

5.0 Avaliação intrínseca e extrínseca

Pode-se concluir, a partir disso tudo, que os sistemas de avaliação são necessariamente embutidos em hierarquias burocráticas. Mas nem isso seria necessariamente o caso. As escolas podem avaliar seu próprio progresso e alterar o percurso segundo seus próprios critérios. Portanto, ainda outra distinção relevante seria se o sistema de avaliação representa um controle intrínseco ou extrínseco.

Quando um sistema de monitoramento é utilizado por um quadro profissional na procura de informações para promover a sua própria melhoria, a iniciativa parte do interior da unidade beneficiada. Nesse caso, a organização

se aproxima de um modelo de associação profissional, onde um dos princípios vigentes seria a auto-regulamentação. O que se avalia, como se avalia e qual ação deveria ser tomada em função dos resultados, representam decisões a serem tomadas através da participação associativa. Já um modelo de monitoramento externo se conforma ao modelo clássico da inspeção burocrática. A determinação do que se avalia, como se avalia e qual ação resulta segue orientações administrativas que se originam em alguma instância acima daquela sendo avaliada.

A dificuldade é que a dependência total dos recursos internos de avaliação, intrínsecos da escola, tende a gerar grandes disparidades nos níveis de desempenho entre escolas. Sem informações comparáveis, não há nenhum mecanismo para avaliar o desempenho dos diversos componentes do sistema e orientar a aplicação dos recursos segundo as prioridades de necessidade. De fato, isto caracteriza a situação do ensino brasileiro nos dias de hoje, onde a falta de qualquer tentativa de avaliação externa obscurece as verdadeiras dimensões das dificuldades educacionais e contribui para a erosão da qualidade do ensino público. Sem uma visão clara das necessidades, é extremamente difícil angariar recursos para o setor e contribuir para a solução dos eventuais problemas. Uma das vantagens da avaliação extrínseca seria, portanto, seu potencial de produzir resultados comparativos e identificar prioridades.

6.0 Comparações em duas dimensões: medidas estáticas e dinâmicas

Os resultados comparativos surgem em duas dimensões distintas. As medidas estáticas do desempenho escolar especificam a posição ou posições das pessoas em escalas de relevância educacional, em algum momento específico. Essas medidas facilitam as comparações transversais entre pessoas ou grupos de pessoas. Medidas dessa natureza são apropriadas para decisões de seleção e de colocação, para verificar o alcance das metas educacionais, para avaliar pessoal e programas, para monitorar entidades organizacionais e incentivar alunos. No entanto, existe um outro tipo de comparação que envolve mudanças de posição ao longo do tempo. Se as medidas estáticas representam um ponto ou pontos, as medidas dinâmicas representam especificações de uma trajetória ou intervalo de pontos ao longo do tempo.

A diferença fundamental pode ser exemplificada através de uma referência às medidas estáticas e dinâmicas na Física. Os pioneiros da Física estiveram limitados à compreensão dos fenômenos estáticos, às propriedades dos objetos em seus estados sem movimento. Mais adiante, os físicos inventaram instrumentos que os ajudaram a compreender os fenômenos dinâmicos, as propriedades dos objetos em movimento. Quando os pioneiros da Física estudaram um peso pendurado numa mola, eles conseguiram medir apenas a extensão da mola quando ela parou de mexer. Mais adiante, os físicos conseguiram compreender tanto a mola quanto o peso como componentes de um sistema em movimento oscilatório.

A distinção entre medidas estáticas e dinâmicas nas ciências físicas é análoga à distinção entre resultado e processo nas Ciências Sociais. Nesse sentido, uma medida de desempenho escolar é uma medida estática, porque o propósito da medida é o de especificar o estado da aprendizagem em momentos particulares do tempo, geralmente no início ou fim do ano letivo. Por outro lado, uma medida do crescimento do desempenho em função do ensino é um exemplo de uma medida dinâmica, porque o propósito da medida é de representar mudanças nos alunos ao longo do tempo.

O Brasil produz e domina uma primeira geração de medidas educacionais, correspondentes às medidas estáticas. Via de regra, os instrumentos e as técnicas de medida permitem ordenar as pessoas em algum momento do tempo, mas não permitem o acompanhamento da mudança em indivíduos ano-a-ano e, a rigor, não são sequer comparáveis série-a-série. A partir de uma perspectiva histórica, essas medidas serviram principalmente a objetivos institucionais, especialmente na seleção dos recursos humanos.

A próxima geração de testes e medidas educacionais deve priorizar a análise dinâmica. Na prática, a análise dinâmica envolve provas de dificuldade variável apropriadas para acompanhar o crescimento cognitivo das pessoas. Exemplos típicos incluem baterias de provas comerciais disponíveis em vários países, elaboradas propositadamente para aplicação nas diversas séries do ensino. Geralmente, as provas são aplicadas de acordo com o nível da série com o propósito de acompanhar o desempenho dos alunos ao longo do ensino fundamental e secundário, série-a-série. Essas provas medem o mesmo conceito mas são elaboradas para variar tanto em conteúdo quanto em dificuldade, muitas vezes por uma margem bastante expressiva. Para acompanhar o crescimento em indivíduos, precisa-se comparar os resultados dessas provas. Isso envolve conjuntos de itens comuns compartilhados entre as provas. Técnicas de calibração são aplicadas às provas assim relacionadas para estabelecer uma única escala curricular comum que permita comparar os resultados.

7.0 Interesses e necessidades: instituições e pessoas

A prioridade que se tem dado no Brasil ao uso dos testes e medidas na seleção dos recursos humanos levanta importantes questões em termos das políticas públicas. Evidentemente, os candidatos a posições universitárias e de emprego deveriam ser considerados na ordem dos resultados da prova. No entanto, essa política tende a restringir as oportunidades dos candidatos menos bem preparados, especialmente a clientela de menor poder aquisitivo, que não tem acesso a oportunidades educacionais plenas. Quando essa política for aplicada sistematicamente ao longo de várias gerações, ela tende a formar uma casta de pobres permanentes que nunca conseguem escapar das circunstâncias de sua pobreza, exacerbando, legitimando e consolidando as disparidades sócio-econômicas existentes.

O pior é que, no Brasil, o processo em si é financiado, não pelas grandes empresas estatais ou pelas universidades federais atendidas, mas pelo próprio candidato, muitos dos quais gastam o que restou de um orçamento familiar bastante limitado para testar sua sorte numa espécie de loteria das vagas públicas. No caso da seleção do recursos humanos para a execução do trabalho de campo do censo demográfico de 1991 do IBGE, as contribuições dos candidatos reprovados excederam em muito à quantia de salários paga aos vencedores do concurso de seleção!

Num primeiro momento, atendeu-se no Brasil apenas o interesse institucional. Num segundo momento, os instrumentos e as técnicas de medida deveriam privilegiar os interesses e as necessidades das *pessoas*, especialmente professores, pais e alunos. Para que isto seja possível, é preciso evoluir a partir das medidas estáticas para as medidas dinâmicas. O uso das medidas educacionais nesse sentido inclui a orientação das pessoas segundo resultados de provas sobre desempenho, aptidão ou interesse, o acompanhamento do progresso em indivíduos e como apoio na tomada de decisões sobre seu ensino. No entanto, a grande maioria dos instrumentos existentes, usados para esses fins, são provas informais administradas individualmente por professores, que não permitem comparações.

As organizações profissionais envolvidas na produção de provas de alta qualidade para fins de seleção de recursos humanos precisam evoluir para a produção de provas de alta qualidade que sirvam às finalidades de monitoramento do processo de aprendizagem em indivíduos. Boas medidas educacionais oferecem informações que representam não apenas a *posição* da pessoa em relação a outras, mas um perfil completo de suas diversas características cognitivas, mais apropriado para acompanhar seu progresso em termos de desempenho e para sugerir as opções mais apropriadas para a evolução de seu conhecimento. O formato das provas deve evoluir a partir da experiência avaliativa classificatória para assumir um compromisso com o acompanhamento do desenvolvimento cognitivo nas pessoas.

8.0 Elenco de testes e medidas: prenúncio de uma nova geração

Ao definir um currículo ou ao desenvolver métodos e materiais de instrução para sala de aula, não seria necessariamente o aluno em particular que deve ser avaliado, mas também o desempenho dos alunos em seu conjunto. Para ser eficiente nesse outro papel, um instrumento de avaliação precisa distinguir múltiplos objetivos curriculares em cada série de ensino. Há vários anos, percebe-se que o instrumento mais adequado para essa finalidade solicita respostas a pequenas amostras de itens representativos de vários domínios de conteúdo, sendo que diferentes alunos respondem a diferentes conjuntos de itens.

Um conjunto maior de itens – geralmente entre 150 e 200 itens – é dividido em blocos distribuídos entre diversas formas da mesma prova de acordo com

um esquema de rodízio que assegura a distribuição equitativa dos blocos em cada sala de aula e a apresentação de cada bloco de itens com cada outro bloco. A amostragem matricial dos itens produz formas alternativas de prova elaboradas para o mesmo nível de dificuldade. No entanto, geralmente não é possível produzir múltiplas formas com exatamente o mesmo nível de dificuldade. Nessas circunstâncias, o emprego de um método para estabelecer a comparabilidade dos resultados se torna imprescindível. Técnicas dos modelos de resposta ao item são aplicadas aos conjuntos de itens comuns para estabelecer a equivalência das formas alternativas, compensando eventuais diferenças de dificuldade. A partir desse momento, as formas se tornam substituíveis entre si e será uma questão de indiferença qual das diversas formas o examinado responde.

A grande vantagem desse esquema é que produz estimativas de desempenho para um elenco de objetivos curriculares bastante abrangente. Para salas de aula e escolas é costumeiro produzir entre 40 e 60 diferentes medidas de objetivos curriculares. Para indivíduos, essas mesmas provas produzem 5 ou 6 medidas do domínio de conteúdos e 2 ou 3 medidas de processos cognitivos, suficientes para caracterizar seu perfil cognitivo. A segurança da prova também é resguardada com esse esquema, porque o examinado não pode prever qual das diversas formas teria que responder e, no caso de um furo na segurança da prova, seria difícil decorar as respostas de todos os 150 e 200 itens que aparecem nas diversas formas da prova. Os mesmos princípios resguardam o uso da prova em ocasiões e datas diferentes, circunstâncias praticamente inevitáveis em certas situações. O rodízio dos blocos gera cinco ou seis formas originais da mesma prova. As formas alternativas permitem que o mesmo aluno seja examinado um igual número de vezes ao longo do ano letivo, sem nunca repetir uma única pergunta. Com a cobertura mais abrangente dos conteúdos e processos cognitivos abordados no conjunto dos cadernos de prova, esse esquema evita a tendência de adaptar o ensino à prova ("ensinar para a prova"), empobrecendo o ensino.

A primeira geração de provas com seus habituais 20 a 30 itens não permite interpretar o significado de um resultado de prova. Simplesmente não há itens em número suficiente para esclarecer esse significado. O maior número de itens obtido com a amostragem matricial permite que os resultados sejam interpretados em termos dos conteúdos e processos cognitivos específicos e genéricos dos itens encontrados em cada ponto da escala. Em vez de uma prova referenciada a normas populacionais, os resultados da escala, obtidos com a amostragem matricial, podem ser compreendidos diretamente em termos do universo de tarefas com eles relacionadas. Basta analisar os conteúdos e processos cognitivos dos itens que se encontram em torno desse mesmo ponto da escala. A amostragem matricial incorpora aos instrumentos e às técnicas de medida possibilidades de análise e de interpretação que esclarecem o verdadeiro

significado dos resultados, resguardando o uso adequado dos instrumentos em situações aplicadas.

O mesmo esquema de rodízio de blocos assegura a comparabilidade dos resultados ano-a-ano. Blocos de itens extraídos das provas antigas devem ser reaproveitados a cada ano na elaboração das provas novas. A partir daí, a cada ano, uma certa porcentagem dos itens testados em anos anteriores seria reaproveitado para assegurar a comparabilidade dos resultados de prova, aplicando os procedimentos de calibração.

Modernas técnicas de impressão permitem a montagem automatizada dos blocos em formas alternativas de prova na hora de sua produção. As provas são impressas frente e verso em duas cores, grampeados na hora da sua impressão, seguidas de uma folha de resposta e empacotados juntos em caixas prontas para distribuição. Quando retiradas das caixas, as formas alternativas de prova já estão intercaladas em ordem, garantindo a justa distribuição dos itens pela sala de aula. Com a velocidade de uma impressora *laser*, a produção dessa nova geração de provas atende 30.000 a 100.000 examinados com prazos e custos muito atraentes.

Resumindo, a nova geração de provas deveria (a) cobrir o conteúdo considerado relevante para os diversos níveis de ensino; (b) fornecer um perfil completo do examinado em termos dos conteúdos e processos cognitivos solicitados; (c) apresentar fidedignidade suficiente para que os resultados obtidos a partir de formas alternativas da mesma prova levem às mesmas conclusões sobre o desempenho dos indivíduos; (d) ser estável o suficiente para permitir comparações de um ano para outro; (e) ter conteúdo comum relacionando provas que tratam da mesma matéria, permitindo que os resultados apresentados nos diversos níveis de ensino sejam apresentados numa escala curricular comum; (f) permitir a apresentação dos resultados numa forma interpretável que todos compreendem facilmente; e ainda (g) ser apresentado na forma de um produto de pronta entrega. Evidentemente esse tipo de medida diferenciada e consistente exige um elevado grau de técnica.

9.0 Provas de lápis-e-papel e modernos bancos de itens

Atualmente, a Fundação Carlos Chagas e a Cesgranrio desenvolvem projetos que visam à produção de provas de lápis-e-papel de alta qualidade para uso na avaliação do ensino. No entanto, a validade do uso desses instrumentos depende do grau de consistência que mantêm com determinado conteúdo curricular e tratamento pedagógico específicos. Para adaptar as provas a uma grande diversidade de currículos e objetivos de avaliação, no futuro, pretende-se colocar à disposição da comunidade modernos bancos de itens implementados no computador. O banco de itens substitui as baterias de provas padronizadas com serviços personalizados que atendem os objetivos específicos de cada um.

A partir dos itens de prova armazenados no banco de itens, novas provas de alta qualidade podem ser elaboradas sob medida em número praticamente

ilimitado. As novas provas podem ser longas ou curtas, difíceis ou fáceis, oferecer uma cobertura compreensiva do currículo ou focalizar apenas algumas áreas mais restritas. Classificações de conteúdo curricular e parâmetros estatísticos catalogados juntos com os gráficos e textos de cada item permitem a escolha automatizada dos itens que melhor atendem os objetivos da prova. A calibração dos itens do banco dispensa qualquer necessidade de um teste piloto prévio com o novo instrumento. Todas suas características técnicas são conhecidas antes de sua aplicação. Esse poder de predição aplica-se tanto às características psicométricas do teste, tais como sua fidedignidade, validade, etc., quanto à calibração dos escores que representam o desempenho dos examinados. Com o reaproveitamento dos itens do banco, esses benefícios redundam em considerável economia de esforço.

Nem por isso pode-se prever todos os usos das provas de papel-e-lápis ou outros instrumentos que venham a serem produzidos a partir dos bancos de item. Constam entre suas potencialidades as finalidades de orientação (ao aconselhar, colocar, promover e certificar alunos), avaliação intrínseca ou extrínseca (ao julgar a suficiência e eficácia dos currículos, métodos pedagógicos e materiais didáticos adotados), administração escolar (ao monitorar o desempenho dos alunos em escolas, programas e delegacias de ensino), consideração das políticas públicas (quando se refere ao desempenho da população na formulação de legislação e da alocação de recursos) e pesquisa (estudos secundários das características que afetam o desempenho do aluno). Na medida em que pode enriquecer a experiência educacional e promover a aprendizagem, essa diversidade seria, naturalmente, bem-vinda.