

# UM SISTEMA NACIONAL DE AVALIAÇÃO DA EDUCAÇÃO: perspectivas atuais

NICIA M. BESSA\*

*"If assessment and instruction can be integrated, education will improve...we can imagine a unification of instructional and assessment design in which both learning and its assessments derive from the same tasks and activities. This would seem to require coordinated theories of learning progress and of diagnostic assessment for adaptive teaching in an instructional domain. Theories of this sort are not now available" (Snow & Mandinach, 1991, p.1.).*

Sistemas educacionais modificam-se, adaptando continuamente objetivos, planos e práticas instrucionais, conforme se altera o contexto em que se inserem - na conceituação de Frederiksen & Collins (1989, p. 27), são "sistemas dinâmicos". Ao propor que a investigação de novas formas de aferir o progresso na aprendizagem, de modo a refletir a influência dos estudos nas áreas da psicologia cognitiva, da teoria instrucional, da psicometria e da tecnologia da informação, Snow & Mandinach (1991) apontam novos caminhos para a adaptação das medidas educacionais às alterações, que já se observam, ou que podem

---

\* Professora da Pontifícia Universidade Católica do Rio de Janeiro - PUC/RJ

ser previstas, nos sistemas educacionais, onde se fazem sentir os efeitos do desenvolvimento sobretudo dos estudos da cognição e das aplicações da informática.

É nessa linha que a proposta do **Seminário Nacional sobre Mídias Educacionais** se coloca, com o objetivo de discutir a renovação do Sistema Nacional de Avaliação da Educação Básica. Este Sistema teve início há alguns anos e, por decisão da atual direção do INEP, merece ser reestudado em suas bases e em sua metodologia, com vistas a uma contínua renovação, que proporcione sua adaptação ao sistema educacional, também em contínua mudança. Iniciando essa nova etapa, o INEP convocou a academia para discutir os fundamentos de uma metodologia de aferição do desempenho escolar.

Ouvir especialistas em medidas educacionais significa abrir espaço para a discussão de novas teorias e de novos métodos. Outros encontros poderão ser organizados para que os responsáveis pelo Sistema de Avaliação conheçam em profundidade os interesses nacionais e regionais, assim como as condições oferecidas por Secretarias de Educação, nos Estados e Municípios, para a implantação de um sistema desse porte. Em um primeiro momento, porém, julgou o INEP ser de importância essencial discutir as bases e as linhas gerais, dentro das quais o Sistema Nacional de Avaliação deva ser renovado.

Assim sendo, foram identificados temas de relevância fundamental para a redefinição do Sistema de Avaliação, principalmente no que se refere à aferição do desempenho dos alunos de 1<sup>a</sup> e de 2<sup>a</sup> graus. Outros temas de caráter metodológico, igualmente relevantes - como conceitos e modelos psicométricos - deverão ser objeto de futuros seminários, já que dependem da discussão de assuntos prioritários.

Ao abrir os trabalhos do Seminário, o Diretor do INEP, Dr. Divonzir Gusso, expôs com detalhe problemas fundamentais que estão preocupando o INEP, ao fazer uma análise do comportamento do sistema educacional brasileiro. Essa introdução permitiu que os objetivos específicos do Seminário se enquadrassem em um conjunto amplo de propósitos, que servem de orientação às ações do órgão na atual gestão.

### **Pesquisas e avaliações apresentadas**

Na parte que precedeu à elaboração dos documentos sobre as cinco questões selecionadas para o debate, os trabalhos apresentados constituíram exemplos concretos, que diziam respeito a problemas teóricos e metodológicos sobre os quais a atenção dos pesquisadores tem convergido nos últimos anos.

Maria Inês Pestana descreve o **Sistema Nacional de Avaliação da Educação Básica (SAEB)**, que a partir de 1988 vem sendo utilizado pelo MEC a fim de avaliar os resultados de políticas e de estratégias educacionais no país, com relação a questões afetas à universalização e à qualidade do ensino. O Sis-

tema de amplo escopo abrange aspectos que vão desde a **gestão educacional até a universalização e a qualidade do ensino**. Os indicadores selecionados pelo Sistema permitem analisar periodicamente relações entre medidas do rendimento dos alunos e variáveis do contexto. Algumas provas para medir o rendimento dos alunos são objetivas, outras de resposta aberta, havendo, inclusive, uma redação para a 5ª e a 7ª séries do 1º grau. O propósito é procurar compreender resultados de políticas e de estratégias adotadas, através de um conjunto de relações de variáveis do contexto social e escolar com resultados de mensurações do rendimento do aluno.

Heraldo M. Vianna descreve as provas utilizadas no segundo International Assessment of Educational Progress (IAEP), as quais incluem itens objetivos e alguns de resposta aberta. Algumas questões, não aplicadas nas amostras brasileiras, requerem que o estudante resolva livremente problemas de Matemática, ou que demonstre seus conhecimentos ao realizar experiências científicas, especialmente montadas para a prova.

Nestes dois projetos observa-se a preocupação em utilizar questões objetivas e questões que têm sido denominadas de "autênticas" - por imitarem as tarefas que os alunos realizam em classe com vistas a adquirirem novos conhecimentos e por serem diretamente comparáveis com os critérios definidos para a avaliação da aprendizagem. O debate entre os especialistas em medidas educacionais continua, atualmente, quanto às vantagens e desvantagens dos dois tipos de aferição do desempenho dos alunos de 1º e de 2º graus. Como resultado positivo desse debate observa-se uma atenção maior dos especialistas ao analisar: a) as características dos dois tipos de questões; b) os possíveis efeitos de um ou de outro tipo de questão, no ensino e na aprendizagem; c) o conceito de validade, face às consequências do emprego de questões dos dois tipos; d) o conceito de fidedignidade, face ao conceito de generalização; e) a complexidade dos processos cognitivos envolvidos ao se responder a questões objetivas ou as chamadas questões "autênticas".

Quanto ao conceito de fidedignidade, Linn, Baker & Dunbar (1991), por exemplo, advertem:

*"The traditional criterion of reliability is subsumed under the transfer and generalizability criterion, but this traditional view of reliability needs to be expanded. Consistency from one part of a test to another, or from one form to another similar (parallel) form is insufficient. Whether conclusions about educational quality are based on scores on fixed-response tests or ratings of performance on written essays, laboratory experiments, or portfolios of student work, the generalization from the specific assessment tasks to the broader domain of achievement needs to be justified" (p.19).*

Não é trivial a tentativa de análise de tarefas propostas ao indivíduo, com vistas a inferir quais os processos cognitivos envolvidos nas respostas observadas. A esse respeito, é pertinente lembrar, com Siegler (1989), que uma análise superficial pode levar a diagnósticos incorretos, já que em suas pesquisas verificou que um mesmo indivíduo pode resolver o mesmo problema usando estratégias e processos diferentes em ocasiões diferentes.

O emprego de questões "autênticas" nem sempre é garantia de que se envolvem processos cognitivos complexos; assim também o emprego de questões objetivas não significa necessariamente o envolvimento exclusivo de conhecimento sobre fatos e o uso de processos cognitivos menos complexos (Linn, Baker & Dunbar, 1991). Na realidade, os processos de análise de questões de ambos os tipos - "task analysis" - têm sido úteis para a compreensão das diferenças e das comunalidades entre as tarefas apresentadas aos alunos e as estratégias que estes usam para cumprí-las. Entretanto, a advertência de Siegler (1989) sobre a dificuldade de se identificar processos empregados pelos indivíduos ao perfazerem a mesma tarefa (ou por um mesmo indivíduo em ocasiões diferentes), com base apenas na análise da tarefa, leva a refletir sobre a subjetividade que permeia as conclusões de tais análises.

De outro lado, com relação às questões "autênticas", a pesquisa psicométrica tem seguidamente demonstrado a dificuldade de se criar sistemas de atribuição de notas (ou conceitos), que sejam fidedignos e, pois, sirvam de base ao diagnóstico, ou à classificação dos alunos em grupos diversos, segundo diferenças observadas em seu desempenho. Em geral, a implantação de um sistema de notas que apresente fidedignidade satisfatória representa um aumento nos custos operacionais. Esse desafio foi enfrentado nos estudos descritos por Inês Pestana, com a adoção de procedimentos que asseguram a fidedignidade das notas atribuídas a redações. E no caso da pesquisa da IAP, constitui uma das razões de não serem aplicadas no Brasil algumas das questões, que consistem em experiências científicas.

Bernardete A. Gatti descreve projeto de avaliação das primeiras séries do 1º grau, em São Paulo, o qual constitui exemplo de soluções referentes a vários temas propostos para reflexão neste Seminário. A preocupação com a integração do sistema de aferição do desempenho escolar com o ensino aparece na análise dos itens das provas, assim como dos erros e acertos dos alunos, realizada pelos próprios professores das escolas em que a avaliação se fez. Esse processo, quando orientado por especialistas, pode levar o professor a refletir sobre a própria prática pedagógica.

Também o emprego de estudos antropológicos, em parceria com a avaliação quantitativa, mostra a preocupação com o conhecimento do contexto em que se desenvolve o aluno e no qual seu desempenho escolar é aferido. Não só o desenvolvimento cognitivo da criança é acompanhado; são estudados o ambiente familiar e da comunidade em que se situa a escola, os processos pedagó-

gicos e a estrutura escolar que envolvem. Tudo isso configura a busca da compreensão dos resultados obtidos pelo sistema escolar.

Ao relatar a avaliação, que se iniciou no 1º grau com o projeto da Secretaria de Educação de Minas Gerais, Heraldo M. Vianna descreve um processo de largo escopo e de longo prazo, no qual a integração do sistema de aferição ao sistema instrucional revela uma faceta importante: a determinação de envolver a comunidade em que se situa a escola na implementação das operações de avaliação, além do corpo docente e discente. É uma tendência interessante, pois evidencia o propósito de estimular a formação de valores e de conceitos, de direcionar sentimentos e motivações com relação à aferição e à avaliação do desempenho do aluno, apresentando-a como parte integrante do processo de ensino e de aprendizagem. Esse é um aspecto que não tem recebido a atenção merecida nos programas de avaliação relativos a grandes populações escolares. Constitui uma iniciativa promissora, na linha defendida por Snow & Farr (1989, p.1):

*The convenient fiction that has long separated theories of cognitive and affective behavior, and caused the conative aspects of behavior to be more or less ignored, must eventually be discarded in the analysis of aptitude, learning and instruction" (p.1).*

Ainda neste programa da Secretaria de Educação de Minas Gerais nota-se a preocupação de procurar compreender os resultados da aferição do desempenho dos alunos dentro do contexto da escola e da cultura local. Já na década de 70, programas desse tipo apresentavam resultados obtidos em testes padronizados de rendimento escolar no contexto de variáveis sociais (Bessa, 1971). No atual projeto de Minas Gerais, indicadores diversos são construídos para formar um conjunto de informações que, relacionadas ao desempenho escolar observado, permitam que se obtenha um diagnóstico da situação educacional do Estado.

Lair Buarque descreve um programa de avaliação do desempenho dos alunos em Língua Portuguesa, com finalidades formativas: o propósito é coletar informações que permitam orientar uma revisão do currículo, da capacitação dos docentes e do processo de alfabetização. A seleção dos instrumentos de medida do desempenho mostra a preocupação também com o aspecto da "autenticidade" - enfatiza-se a interpretação e a produção de texto, sendo que em algumas séries a criança escreve uma "história". Ainda neste caso, transparece a preocupação em formar escalas, que preservem a confiabilidade dos resultados.

Com crianças das primeiras séries do 1º grau, nem sempre o emprego de provas de resposta livre prende-se a razões de validade aparente ("face validity"), mas sim ao fato de ser inviável, na prática, o uso de provas objetivas. Ob-

serva-se na literatura especializada, entretanto, que alguns pretendem justificar o emprego generalizado de provas "autênticas" por sua validade aparente. Outros, com maior rigor metodológico, exigem a aplicação de diversos critérios de validação, entre os quais o da generalização, que parte do desempenho nas provas para se aplicar ao desempenho nos trabalhos escolares regulares. Por exemplo, Linn, Baker & Dunbar (1991) concluem que não se pode deixar de justificar, com base nas pesquisas empíricas, a generalização de conclusões sobre resultados obtidos em provas de rendimento escolar, com relação ao desempenho do aluno em tarefas de outra natureza.

No mesmo sentido se expressa Lorrie Shepard (1991), ao fazer a defesa das provas "autênticas":

*... the new alternative assessments must be subjected to construct investigations designed to teste the generality of conclusions about student's knowledge and reasoning ability on the bases of what he or she is able to do in the assessment context" (p.10).*

### **As conclusões dos cinco GRUPOS**

As tendências que caracterizam as propostas expostas pelos pesquisadores, quanto à fundamentação teórica da aferição do rendimento dos alunos na escola, são as mesmas observadas nos relatórios sobre os cinco temas apresentados à discussão no Seminário.

As conclusões do GRUPO 1 mostram, basicamente, duas preocupações: a) com o acompanhamento do desenvolvimento cognitivo do aluno; b) com a preparação de pessoal, de modo a se garantir a qualidade e a propriedade na elaboração de instrumentos e na interpretação dos resultados observados. O GRUPO 1 reconhece que existem diferentes tendências, na prática, quanto a teorias da aprendizagem "implícitas" (Shepard, 1991, p.3.) na elaboração de instrumentos para aferição do rendimento escolar, ou na descrição do desenvolvimento cognitivo do aluno. São concepções que, na análise do Cole (1990), espelham vertentes das teorias behavioristas ou das teorias cognitivistas.

As duas grandes preocupações demonstradas no relatório do GRUPO 1 parecem em linha com a posição de Snow & Mandinach (1991, p.10), quando argumentam contra a base lógica das taxonomias das especificações que orientam a elaboração de questões de provas, por não representarem, a seu ver, as estruturas e processos psicológicos envolvidos no desempenho do aluno nas tarefas escolares. De outro lado, as propostas do GRUPO 1 parecem reconhecer a necessidade de certa cautela, principalmente quando focalizam a formação de pessoal e preconizam a realização de pesquisas e estudos. Também, nesse ponto, afinam-se as sugestões do GRUPO 1 com o pensamento de Snow & Mandinach (1991), quando estes observam que apenas se inicia a proposta de teorias

e modelos psicométricos que fundamentem a medida educacional em linha com concepções do progresso escolar, compreendido no âmbito do desenvolvimento dos estudos da cognição e da afetividade:

*"The next wave of theory and research on educational assessment will come from recognition that interpretations of achievement measures rest on psychological constructs and have to be validated as such ... Developments in cognitive and conative psychology suggest new ways of thinking about the constructs underlying educational measures, new constructs to be assessed, and new kinds of test designs" (p.10).*

Se, de um lado, esse é um caminho ainda difícil e incerto (Bennett, 1991; Cole, 1990; Shepard, 1991; Diegler, 1989), de outro lado, há já exemplo de modelos psicométricos com base em teorias da cognição e na psicologia educacional, tendo em vista uma integração com as teorias e modelos instrucionais (Mislevy, 1989, Mislevy & Yamamoto, 1991; Sheehan & Mislevy, 1990).

É nesse conjunto de preocupações que as recomendações dos GRUPOS 1 e 5 se completam, ao enfatizarem a necessidade de conscientização, de formação e de aperfeiçoamento de pessoal com relação ao processo de avaliação educacional.

O GRUPO 2 também focaliza a importância da formação do avaliador, em várias instâncias educacionais e sugere um profundo envolvimento de todos os interessados que fazem parte do contexto em que se realiza a avaliação. Para o GRUPO 2, não se trata apenas de integrar as tarefas propostas ao aluno, com o fim de aferir seu rendimento escolar, às tarefas que pretendem facilitar a aprendizagem; mais do que isso, trata-se de fundamentar a construção de instrumentos e a seleção de indicadores no estudo das experiências, preocupações e interesses de quantos estão envolvidos no processo de avaliação.

Não é nova a discussão da proposta de que instrumentos para aferição do rendimento escolar sejam construídos de modo a refletir o contexto educacional e cultural específico em que deverão ser usados. Na realidade, psicometristas como Lord (1980) e outros desenvolveram teorias e estratégias para o caso extremo de adaptar a prova a cada indivíduo ("tailored tests"). As questões que a proposta suscita dizem respeito, principalmente, à comparabilidade de resultados de provas diferentes, quando aplicadas a diferentes grupos de indivíduos (Linn, Baker & Dunbar, 1991; Miller-Jones, 1989), assim como à validade do construto (Laosa, 1991).

As recomendações do GRUPO 2 situam-se na vertente das propostas de Snow & Mandinach (1991), ao dar relevo a aspectos afetivos na integração da avaliação com os processos instrucionais. E vai mais longe, ao preconizar um envolvimento de todos os interessado nos resultados da avaliação; nessa linha,

ênfatisa o respeito e a valorização da pessoa de cada um dos participantes do processo. Pode-se perguntar se não terá sido a falta de envolvimento dos interessados na avaliação uma falha que contribuiu, na história das aplicações das medidas educacionais, para estimular, em muitos setores de vários sistemas educacionais, a oposição à aferição do rendimento dos alunos nas escolas de nível básico e secundário.

O GRUPO 3 ênfatisa, mais uma vez, a necessidade de formação e de aperfeiçoamento de pessoal na área da avaliação e, em especial, na área das medidas educacionais. Também neste GRUPO, o contexto em que se realiza o trabalho escolar é compreendido em sentido da maior amplitude, abrangendo a escola, a família e a comunidade.

A pesquisa qualitativa, em educação, tem partido de diferentes pressupostos metodológicos (Jacob, 1989), porém a utilização dos resultados obtidos tem sido comumente a de melhor compreender o conjunto de relações percebidas no contexto de um sistema educacional. De acordo com recente mapeamento feito por Greene, Caracelli & Graham (1989), a utilização, em um mesmo estudo, de métodos qualitativos e quantitativos se apresenta sob várias formas e parte de diversos pressupostos teóricos. Do ponto de vista das bases teóricas, o mapeamento de Greene *et alii* pode ser reduzido, em última análise, a dois grupos principais: a) o dos que negam a compatibilidade dos pressupostos dos métodos qualitativos e quantitativos de pesquisa e avaliação educacional; b) o dos que aceitam a utilização conjunta de métodos qualitativos e quantitativos, com base no conceito de triangulação, entendida como o uso de múltiplos métodos no estudo de um mesmo fenômeno com vistas a fortalecer a validade dos resultados. Entre os pesquisadores que misturam as duas metodologias, seja com base teórica, seja por motivos práticos, variam as finalidades com que o fazem: a) alguns pretendem apenas garantir validade aos resultados através da triangulação; b) há "designs" que visam a complementar os resultados obtidos com uma das metodologias com os resultados obtidos com o uso de outras, ao estudar diferentes facetas dos fenômenos em foco; c) há "designs" em que a combinação é usada com o fim de desenvolver o estudo em que as duas metodologias são usadas sequencialmente, os resultados de uma servindo de base à outra na seleção de processos e de instrumentos; d) alguns advogam a utilização de diferentes bases epistemológicas em um mesmo "design", com o propósito de descobrir resultados contraditórios; e) há "designs" em que o propósito é expandir os objetivos, ou a amplitude dos resultados, sendo a duplicidade metodológica usada ao se investigar diversos fenômenos.

Partindo do mapeamento de Greene *et alii*, pode-se interpretar a posição do GRUPO 3 como melhor alinhada com os pontos de vista daqueles que pretendem utilizar metodologias qualitativas e quantitativas, em um mesmo estudo, com a finalidade de proporcionar meios de desenvolver a investigação, ou de complementar resultados, ou de expandir o escopo da avaliação. Sem descu-

rar do rigor metodológico, o GRUPO 3 não vê, na mistura de métodos, o propósito de triangulação exclusivamente, mas também a melhor compreensão dos resultados - ou seja, o desvelamento de facetas e de diversos fenômenos, de modo a identificar condições, dimensões, indicadores que orientem a investigação.

O GRUPO 4 se posiciona, em primeiro lugar, a favor de sistemas regionais de avaliação. A preferência por sistemas regionais vem de encontro à tendência atual de fortalecimento dos agentes locais que atuam na educação de 1ª e de 2ª graus. De outro lado, o GRUPO 4 reconhece a possibilidade de coexistência de um sistema nacional de avaliação do desempenho escolar e de sistemas regionais.

Os problemas que acompanham as propostas de sistemas nacionais ou de sistemas regionais aparecem mais agudamente em países onde a heterogeneidade cultural dificulta a construção de instrumentos de medida cujos resultados sejam aceitos como válidos e generalizáveis em âmbito nacional. De um lado, a tendência à contextualização da aferição do rendimento escolar exprime a necessidade de compreender o desempenho dentro da situação em que é observado. De outro lado, a necessidade de obter informações em âmbito nacional de modo a fundamentar políticas educacionais de abrangência geral leva à procura de indicadores comuns a diversas regiões. O problema da comparabilidade dos resultados persiste, entretanto, dado que, em contextos culturais muito diferentes, esses resultados podem ter um sentido também diferente.

O caso do programa americano de âmbito nacional - **National Assessment of Educational Progress (NAEP)** - é ilustrativo: mesmo em um país em que não há currículo obrigatório, por lei, em âmbito nacional, NAEP tem conseguido oferecer um panorama geral, em nível macrocópico, dos resultados obtidos em escolas norte-americanas. Deve-se notar, porém, que o programa não produz resultados por aluno, mas apenas agregados por todo o país. A proposta do NAEP vir a oferecer resultados agregados por Estado tem sido discutida não só quanto às possíveis vantagens ou desvantagens do ponto de vista de política educacional, como também quanto ao sentido que se possa dar a interpretação das comparações entre Estados (Ferrara & Thornton, 1988).

O GRUPO 4 preconiza prudência: preconiza um trabalho por etapas, para dar tempo ao amadurecimento de conceitos e de processos; preconiza, por fim, um trabalho que, centralizado no INEP, fortaleça órgãos regionais de avaliação.

Ao focalizar as questões de formação e aperfeiçoamento de pessoal e a formação de núcleos de pesquisa e desenvolvimento em diferentes regiões do país, o GRUPO 5 apoia explicitamente o envolvimento não só de especialistas em medidas educacionais como a conscientização de diversos agentes do processo de avaliação e dos órgãos das comunidades, interessados na educação. Ainda neste ponto, portanto, transparece a preocupação, que permeia os relatórios

dos demais GRUPOS, com o envolvimento das comunidades numa demonstração de que a tendência à integração da avaliação com o processo instrucional é entendida em seu sentido mais amplo, abrangendo o ambiente familiar e comunitário. Para que o processo seja efetivo, entende o GRUPO 5 que a conscientização e a informação adequada deve ser levada a todos os interessados no processo.

De outro lado, o processo inverso é apontado como fundamental: núcleos regionais - que, na situação atual do sistema educacional brasileiro, necessitam apoio do INEP para sua formação e manutenção - devem ser a fonte de informação para o centro nacional, diretamente a eles ligados por rede de computadores. É clara a tendência à descentralização, intimamente associada a um processo de integração com o INEP, ao qual se atribui um papel de agente dessa integração e de fomento da avaliação, da formação e do aperfeiçoamento de pessoal. A contextualização da informação é privilegiada, ao mesmo tempo que não se exclui a integração das informações em âmbito nacional por um núcleo central - no caso, o INEP.

Nota-se, pois, na proposta do GRUPO 5, duas preocupações fundamentais: a) o aproveitamento das fontes de pesquisa e avaliação já existentes no país e a integração das informações obtidas; b) o reflexo - até agora minimamente observado - dos resultados da avaliação de alunos e do próprio sistema educacional sobre os conhecimentos, os valores, as atitudes, os modos de agir e de sentir de quantos se envolvem com o ensino e a aprendizagem nas escolas de 1ª e 2ª graus. A aferição do rendimento do aluno é concebida como parte da prática pedagógica e como parte de um trabalho de avaliação, que tem o caráter de uma pesquisa educacional.

Referindo-se exclusivamente à aferição do rendimento escolar Frederiksen & Collins (1989) tomam uma posição radical, em que entendem a concepção de validade dos resultados obtidos em provas escolares de modo a ter em consideração os efeitos de mudanças no processo instrucional introduzidas por influência das informações obtidas através de medidas do rendimento escolar. Neste sentido, concebem "validade sistêmica" de resultados obtidos em medidas educacionais como observada quando, por influência dos mesmos, são introduzidas alterações no sistema educacional, as quais favorecem o desenvolvimento cognitivo do aluno.

A posição de Frederiksen & Collins é polêmica, seja por presumir que resultados da aferição do rendimento escolar possam ter reflexos visíveis e profundos no sistema educacional, seja por ignorar os problemas suscitados pela pretensão de partir de tais resultados para fazer inferências sobre processo cognitivos.

De outro lado, verifica-se que o GRUPO 5, dentro dos limites de realismo, que transparecem em suas sugestões, aponta caminhos a serem percorridos em prazo mais longo ou mais curto - todos certamente com o objetivo seme-

lhante ao de Frederiksen & Collins, qual seja, o de fazer da informação um instrumento para se conseguir introduzir alterações nos processos com que se pretende facilitar as condições de aquisição de conhecimento nas escolas brasileiras. Embora menos radical, a proposta do GRUPO 5 é tão ambiciosa - conquanto bastante realista - quanto a de Frederiksen & Collins, pois em ambos os casos é dado o maior relevo ao papel formativo da avaliação.

A preocupação com a utilização dos resultados de provas de rendimento escolar com propósitos formativos tem sido mais objeto de retórica do que de iniciativas de ordem prática no Brasil e em outros países. Ao concluir a análise da proposta de um sistema nacional de aferição do rendimento escolar na Inglaterra, Nuttal (1990) expressa seus receios:

*"It is usually the summative and evaluative purposes that predominate, often distorting the curriculum and learning methods, while the formative purpose gets down-graded or even lost. This is the fear of many in relation to the new national assessment system being introduced in England and Wales" (p.380).*

A aferição apenas para aferir pode ter reflexos mais negativos do que positivos em educação. Os cinco GRUPOS, que discutiram e apresentaram suas sugestões neste Seminário, apontam caminhos para implementar um sistema de aferição e de avaliação que tenha condições de servir ao aluno, em última análise. A responsabilidade de implementação de um tal sistema não pode caber exclusivamente a um núcleo central - no caso, o INEP, cabe, igualmente, a cada Estado, a cada escola, a cada centro de pesquisa e a cada agente do processo de avaliação.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BENNET, R. E. (1991) On the meaning of constructed response. **Research Report -91-63-Educational Testing Service**. Princeton, N.J.: Educational Testing Service.
- BESSA, N. M. (1971) **Alunos do curso colegial - Planos e características socioeconômicas**. R.J.: Fundação Getúlio Vargas.
- COLE, N.S. (1990) Conceptions of educational achievement. **Educational Researcher**, 19(3), 2-7.
- FERRARA, S. F. & Thornton, S.J. (1988) Using NAEP for interstate comparisons: The beginnings of a "National Achievement Test" and "National Curriculum". **Educational Evaluation and Policy Analysis**, 10(3), 200-211.
- FREDERIKSEN, J.R. & Collins, A. (1989). A system approach to educational testing. **Educational Researcher**, 18(9), 27-32.
- GREENE, J.C., Caracelli, V. J. & Graham, W. (1989) Toward a conceptual framework for mixed-method evaluation designs. **Educational Evaluation and Policy Analysis**, 11(3), 255-274.
- JACOB, E. (1989). Qualitative research: "A defense of traditions". **Review of Educational Research**. 59(2), 229-235.

- LAOSA, L.M. The cultural context of construct validity, and the ethics of generalizability. **Research Report-91-51- Educational Testing Service.** Princeton, N. J. Educational Testing Service.
- LINN, R.L., Baker, E. L. & Dunbar, S.B. (1991) Complex performance-based assessment: Expectations and validation criterion. **Educational Researcher**, 20(8), 15-21.
- LORD, F.M. (1980) Applications of item response theory to practical testing problems. Hillsdale, N.J.: Lawrence Erlbaum.
- MILLER-Jones, D. (1989) Culture and testing. **American Psychologist**, (1989) 44, 360-366.
- MISLEVY, R.J. (1989) Foundations of a new test theory. **Research Report-89-52- Educational Testing Service.** Princeton, N.J.: Educational Testing Service.
- MISLEVY, R. J., Yamamoto, K & Anaker, S. (1991) Toward a test theory for assessing student understanding. **Research Report-91-32-O NR - Educational Testing Service.** Princeton, N.J.: Educational Testing Service.
- NUTTAL, D.L. (1990) Proposals for a National System of Assessment in England and Wales. **International Journal of Education Research**, 14(4), 373-381.
- SHEEHAN, K. & Mislevy, R.J. (1990) Integrating cognitive and psychometric models to measure document literacy, **Journal of Educational Measurement**, 27(3), 255-272.
- SHERPAR, L.A. (1991) Psychometrician's beliefs about learning. **Educational Researcher**, 20(6), 2-16.
- SIEGLER, R.S. —) Strategy diversity and cognitive assessment., **Educational Researcher** 18(9), 15-20.
- SNOW, R.E. & Farr, M.J. (Eds.) (1987) Cognitive - conative - affective processes in aptitude, learning and instruction : An introduction. In R.E. Snow & M.J.Farr (Eds.), **Aptitude Learning and Instruction**, vol.3: Cognitive and Affective Process Analyses. Hillsdale, N.J. : Lawrence Erlbaum, 1-10.
- SHOW, R.E. & Mandinach, E.B. (1991) Integrating assessment and instruction: A research and development agenda. **Research Report-91-8- Educational Testing Service.** Princeton, N. J. Educational Testing Service.



## **Apresentação**

### **A Escola Pública de 1º grau: limites e possibilidades**

Aida Maria Monteiro Silva

### **A Escola Pública: um estudo sobre a aprendizagem nas séries iniciais**

Cláudia Davis, Yara Lúcia Esposito

### **Qualidade da Educação debaixo do tapete**

Sérgio Costa Ribeiro

### **Avaliando a avaliação - da prática à pesquisa**

Heraldo Marelím Vianna

## **SEMINÁRIO NACIONAL SOBRE MEDIDAS EDUCACIONAIS**

### **Medida e Avaliação Educacional: problemas e tendências**

Divonzir Arthur Gusso

### **Seminário Nacional sobre Medidas Educacionais: temas para discussão**

Nícia M. Bessa

### **O Sistema Nacional de Avaliação da Educação Básica**

Maria Inês Gomes de Sá Pestana

### **Avaliação da Jornada Única em São Paulo**

Bernardete A. Gatti

### **Avaliação do Ciclo Básico de Alfabetização em Minas Gerais**

Heraldo Marelím Vianna

### **Avaliação do Desempenho da Rede Pública Escolar do Estado de Pernambuco na Área de Linguagem**

Lair L. Buarque, Zélia M. Hígino, Elisabete M. de Miranda, Maria Helena Du-beux e Ivane Pedrosa

### **Avaliação do Desempenho em Matemática e Ciências - uma experiência em São Paulo e em Fortaleza**

Heraldo Marelím Vianna

### **Discussão de Temas sobre Avaliação - manifestação de pesquisadores e avaliadores**

### **Um Sistema Nacional de Avaliação da Educação - perspectivas atuais**

Nícia M. Bessa