

A PRÁTICA DA AVALIAÇÃO EDUCACIONAL: ALGUMAS COLOCAÇÕES METODOLÓGICAS¹

A avaliação educacional nos dias de hoje atrai a atenção de pesquisadores face à necessidade de promover o levantamento de um amplo espectro de elementos que possibilitem a análise dos sistemas educacionais em suas várias dimensões. Esse processo, entretanto, atravessa momento crítico, porque, conforme Stufflebeam *et al.* (1971), não possui uma teoria perfeitamente estruturada que traduza um consenso entre educadores. Além do mais, inexistente uma tipologia de informações fundamentais para o processo decisório, há insuficiência de instrumentos e planejamentos adequados aos diversos fenômenos educacionais e, ainda, destaca-se a falta de um sistema que possibilite a organização, o processamento e o relatório de informações necessárias à avaliação educacional. O trabalho de avaliação exige, conseqüentemente, certa criatividade e independe da adoção de modelos desenvolvidos em outros contextos. Assim, a pesquisa em avaliação educacional, inclusive quando restrita ao rendimento escolar, exige definições e procedimentos *ad hoc*, para que possa ser realizada com pleno êxito. É evidente que, não havendo um consenso difícil, às vezes, entre educadores ocorram, especialmente nessa área, incompreensões quase

¹ Artigo publicado em *Cadernos de Pesquisa*, n. 69, p. 40-47, maio 1989.

sempre centradas em aspectos não necessariamente relevantes para os propósitos da avaliação.

É necessário ressaltar que, ao contrário da pesquisa, a avaliação educacional não visa à generalização dos resultados com vistas ao estabelecimento de princípios ou de leis. A avaliação tem por objetivo gerar conhecimentos que levem a decisões que tenham consequências imediatas na prática educacional. Desse modo, no caso particular do rendimento escolar, por exemplo, a avaliação procura obter dados que possibilitem a tomada de decisões relativas ao desempenho individual e/ou coletivo em face de um determinado programa curricular. A partir de elementos empiricamente coletados, a avaliação procura descrever, da melhor forma possível, o fenômeno considerado, para possibilitar a fundamentação do processo decisório com base em dados da realidade.

A avaliação educacional, como área de investigação científica é recente, e as posições teóricas sobre as prioridades a considerar contribuíram para a formulação de diferentes definições e a criação de grande variedade de modelos. O problema primeiro está em definir, com relativa precisão, duas palavras frequentemente usadas de forma intercambiável: medir e avaliar.

Medir é uma operação de quantificação, em que se atribuem valores numéricos, segundo critérios preestabelecidos, a características dos indivíduos, para estabelecer o quanto possuem das mesmas. O índice quantitativo, obtido por intermédio da medida, identifica o status do indivíduo face à característica. Com referência à avaliação, a medida é um passo inicial, bastante importante, às vezes, mas não é condição necessária, nem suficiente, para que a avaliação se efetue. Eventualmente, a medida pode levar à avaliação que, entretanto, só se realiza quando são expressos julgamentos de valor.

Avaliar é determinar o valor de alguma coisa para um determinado fim. A avaliação educacional visa, pois, à coleta de informações para julgar o valor de um programa, produto, procedimento ou objetivo (WORTHEN E SANDERS, 1973) ou, ainda, a apreciar a utilidade potencial de abordagens alternativas para atingir determinados propósitos. A avaliação refere-se, assim, a atividades sistemáticas ou formais para o estabelecimento do valor de fenômenos educacionais (POPHAM, 1975), quaisquer que sejam.

A avaliação, para alguns, é um processo assistemático baseado na opinião de um especialista, é um julgamento emitido por um profissional. São comuns, na área educacional, “avaliações” informais para a tomada de certas decisões. Um livro é adotado em vez de outro, uma metodologia de ensino é empregada em substituição à outra, apenas com base em avaliações assistemáticas e impressionistas. A chamada “avaliação”, nesses casos, limita-se a uma escolha, com base em percepções, da que seria a melhor alternativa. É, pois, uma simples opção, sem fundamento científico. A avaliação, ao contrário, decorre de um esforço sistemático para definição de critérios, em função dos quais se coletam informações precisas para julgar o valor de cada alternativa apresentada. Avaliar é, assim, emitir um julgamento de valor sobre a característica focalizada, podendo esse valor basear-se parcial, mas não exclusivamente, em dados quantitativos.

A Tyler (1942) coube a difusão da definição de avaliação como um processo de comparação entre os dados do desempenho e os objetivos instrucionais preestabelecidos. Essa definição desfrutou de grande receptividade nos meios educacionais e, com pequenas variações, foi incorporada a alguns modelos teóricos como, por exemplo, o de Hammond (s.d.) e o de Metfessel e Michael (1967).

Stufflebeam *et al.* (1971) desenvolveram um modelo centralizado na ideia de que a avaliação deve permitir aos administradores a tomada de decisões e, coerentemente, definiram avaliação como o processo de identificar e coletar informações que permitam decidir entre várias alternativas.

Outros teóricos, juntamente com as suas contribuições de estratégias para a investigação avaliativa, propuseram também definições que, em maior ou menor grau, são aceitas por muitos praticantes da avaliação educacional. Entre essas definições, destacam-se a de Provus (1971), que apresenta a avaliação como um processo de comparação entre desempenho e padrões, e a de Stake (1967), que a caracteriza como descrição e julgamento de programas educacionais.

A avaliação, como campo emergente na área educacional, tem recebido contribuições provenientes de várias fontes, entre as quais destacam-se as de Michael Scriven, que marcaram, profundamente, a teoria da avaliação educacional. Scriven (1967) concebe esse processo como um levantamento sistemático de informações e sua posterior análise para fins de determinar o

valor de um fenômeno educacional. Essa definição, centralizada no problema do valor, influenciou o pensamento de grande parte dos teóricos e, praticamente, dos estudiosos da avaliação moderna, inclusive de alguns que não se preocuparam em detalhar e explicitar a questão, como foi o caso de Stufflebeam. Analisada a conceituação estabelecida por esse último, verifica-se que ela também incluiu um julgamento de valor, ainda que não o tenha explicitado. Escolher essa ou aquela alternativa, isto é, decidir, conforme estabeleceu Stufflebeam, é julgar o valor de uma ou de outra alternativa, optando pela melhor. Assim, na definição de Stufflebeam *et al.* (1971), está implícito, também, um julgamento de valor (VIANNA, 1982a).

A avaliação no contexto educacional é uma necessidade imperativa e exige uma metodologia que possibilite a coleta de informações para decisões fundamentadas. Um esquema de planejamento frequentemente encontrado em projetos de avaliação é o baseado na análise das diferenças apresentadas antes e após o tratamento instrucional. Essa estratégia, ainda que útil em certas condições, nem sempre fornece informações detalhadas que permitam tomar decisões complexas. Outra opção estratégica, também amplamente utilizada em avaliação, é a do planejamento experimental, que caracteriza a pesquisa empírica, mas que nem sempre é suficientemente eficaz para a avaliação de alguns fenômenos educacionais, tendo em vista a circunstância de que a avaliação se processa num quadro natural, em que as situações nem sempre são bem estruturadas e, por isso, tornam-se difíceis as condições de controle exigidas pelo planejamento experimental. Assim, tendo em vista essa problemática, vários especialistas procuraram desenvolver novas estratégias para dar à avaliação um sentido mais eficaz.

As diferenças existentes entre os modelos decorrem do fato de estabelecerem prioridades diversas para os problemas de avaliação educacional. Assim, como exemplificação, sem aprofundar a análise de todos os modelos anteriormente mencionados, observa-se que Tyler (1942) se encontra na problemática da convergência entre desempenhos e objetivos instrucionais, Stake (1967) baseia-se na análise de variáveis antecedentes, intermediárias (*transactions*) e resultantes, Stufflebeam *et al.* (1971), através do exame do contexto, entrada (*input*), processo e produto, visa a obter informações que permitam a tomada de decisões pelos administradores.

O avaliador educacional, ao selecionar determinado modelo teórico, a fim de desenvolver um projeto, deverá levar em consideração a natureza do problema a investigar, os recursos disponíveis e sua própria situação pessoal. Os modelos não se propõem a resolver todos os problemas que se apresentem ao avaliador; objetivam, na verdade, permitir que o avaliador dimensione, adequadamente, os seus projetos, para evitar que deficiências de planejamento invalidem o processo e levem a falsas decisões (VIANNA, 1982a).

Cronbach, no artigo *Course improvement through evaluation* (1963), ainda que não apresente um modelo de avaliação, oferece um conjunto de ideias altamente provocadoras que tiveram grande impacto na década de 60, influenciando trabalhos como o de Scriven (1967) e o de Stake (1967) que, por sua vez, repercutiram profundamente na prática de avaliação educacional. O ensaio de Cronbach é de grande valor e discute, sobretudo, quatro pontos: a associação entre avaliação e processo de tomada de decisão, os diferentes papéis da avaliação educacional, o desempenho do estudante como critério de avaliação de cursos, algumas técnicas de medidas à disposição do avaliador educacional.

A avaliação, no seu sentido mais amplo, pode ser definida como um processo que visa à coleta e ao uso de informações que permitam tomar decisões sobre um programa educacional. A avaliação, portanto, segundo Cronbach (1963), deve ser entendida como uma entidade diversificada, que exige a tomada de diversos tipos de decisões e o uso de uma grande variedade de informações. A avaliação com vistas ao aprimoramento de currículos não deve ser confundida, como muitos o fazem, com a construção de instrumentos de medida e a obtenção de escores fidedignos, processos que, eventualmente, podem entrar no contexto da avaliação mas que não são indispensáveis para que ela possa atingir os seus objetivos.

Cronbach mostra que a avaliação é usada com o objetivo de tomar alguns tipos de decisões, entre os quais:

1. determinar se os métodos de ensino e o material instrucional utilizado no desenvolvimento de um programa são, realmente, eficientes;
2. identificar as necessidades dos alunos para possibilitar o planejamento da instrução, julgando o mérito dos es-

tudantes para fins de seleção e agrupamento e fazendo com que conheçam seu progresso e suas deficiências;

3. julgar a eficiência do sistema de ensino e dos professores;
4. Assim, de acordo com Cronbach, no primeiro caso, a avaliação permitiria decisões que levariam ao aperfeiçoamento do currículo; no segundo, referir-se-ia aos alunos submetidos a determinado programa; finalmente, no terceiro e último caso, as decisões seriam de natureza administrativa.

Cronbach discute, particularmente, alguns pontos fundamentais que não podem ser ignorados pelo avaliador educacional. Inicialmente, enfatiza o seguinte: “quando a avaliação visa ao aprimoramento de cursos, seu principal objetivo é verificar quais os efeitos do curso, ou seja, quais as mudanças que produz no estudante”. O problema, segundo a perspectiva desse estudioso, não está em determinar se um curso é eficiente ou ineficiente apenas. É preciso lembrar que os resultados da instrução são multidimensionais e, desse modo, a avaliação deve promover o mapeamento de todos os efeitos do curso em cada uma de suas dimensões. Um erro frequente está na concentração em um único escore de diversos desempenhos esperados após a realização de um curso. Isso pode ser enganador, pois um insucesso numa dimensão pode ser compensado pelo sucesso em outra. Os escores compostos englobam e, muitas vezes, ocultam julgamentos sobre a importância de vários resultados; desse modo, para fins de avaliação, é importante que os resultados sejam apresentados separadamente, a fim de que se tenha uma ideia real das mudanças que estariam ocorrendo no estudante como decorrência da influência exercida pelo currículo.

“A avaliação presta um grande serviço quando identifica os aspectos do curso que necessitam de revisão”, no dizer de Cronbach (1963). É evidente que qualquer especialista em currículo gostaria de apresentar evidências sobre a eficiência de seu produto; entretanto, conforme observa Cronbach, esses especialistas costumam relutar na aceitação de uma avaliação externa. O procedimento habitual consiste em submeter o produto à avaliação somente depois de terminado, com vistas a uma confirmação do que foi previamente estabelecido. Esse comporta-

mento não proporcionará bons resultados, além de traduzir um menosprezo pelo importante papel que o avaliador pode, efetivamente, desempenhar. A fim de apresentar um papel influente, a avaliação deve acompanhar o desenvolvimento do currículo, quando o mesmo ainda se acha em estado fluido, para usar palavras de Cronbach. A avaliação, assim entendida, possibilita a criação de conhecimentos sobre a natureza das capacidades que constituem os objetivos educacionais do projeto.

“A comparação de cursos não deve ser o objetivo dominante da avaliação”, na concepção de Cronbach. A comparação de resultados de avaliações de cursos deve ser cautelosa, para evitar que decisões errôneas sejam tomadas. As diferenças entre os escores médios de diferentes cursos, geralmente, costumam ser pequenas, em virtude da grande diferença que há entre e intra grupos submetidos ao mesmo curso. A impossibilidade de equiparar (*equate*) diferentes grupos prejudica a interpretação dos resultados e representa um problema difícil de superar nos estudos comparativos de cursos. Além do mais, ainda conforme Cronbach, em experimentos educacionais, é difícil ocultar dos estudantes o fato de que integram um grupo experimental, sendo igualmente complexo o controle do viés dos professores numa situação experimental. Ocorrem mudanças comportamentais e, assim, nem sempre se pode afirmar, com convicção, que um determinado resultado decorre, efetivamente, da própria inovação ou do fato de alunos e professores terem sido colocados diante de uma situação experimental.

A comparação entre grupos pode oferecer resultados equívocos; desse modo, Cronbach (1963) propõe que estudos formais sejam planejados, sobretudo, para determinar o desempenho, após o curso, de um grupo perfeitamente conhecido, a fim de verificar objetivos importantes e ocorrência de efeitos colaterais. Cronbach chama a atenção para o fato de que, em um experimento onde os tratamentos comparados diferem em inúmeros aspectos, a ocorrência de uma pequena diferença numérica em favor da situação nova não significa grande coisa e não contribui para o aumento de nossos conhecimentos. Estudos analíticos, realizados em escala menor mas em condições controladas, de alternativas de um mesmo curso, oferecem melhores resultados do que pesquisas de campo que aplicam tratamentos dissimilares a grupos diferentes.

Cronbach destaca o fato de considerar mais importante os dados relativos a um item do que os escores do teste. O escore global pode dar confiança ou não em relação a um determinado currículo, mas pouco informa como aprimorá-lo. Outro aspecto importante em avaliação de currículo, conforme mostra Cronbach, refere-se à ajustagem do instrumento de medida ao currículo, que não deve ser motivo de preocupação para o avaliador, ainda que possa parecer surpreendente, porque, em outras situações, isso não deve ocorrer. Numa avaliação ideal de currículo, todos os tipos desejáveis de proficiência devem ser medidos e não apenas aqueles referentes a alguns objetivos selecionados pelo construtor do currículo. Se há interesse em saber se o currículo alcança os seus objetivos, impõe-se o ajustamento do teste a ele, mas, se a intenção é a de verificar até que ponto o currículo atende a interesses mais amplos, o ajustamento não é necessário, tendo em vista o desejo de que sejam verificados todos os objetivos possíveis.

Outro ponto enfocado por Cronbach diz respeito à distinção entre testes factuais e testes para verificar processos mentais complexos, segundo a terminologia de Bloom (1956). A classificação dos itens de acordo com as categorias propostas por Bloom é difícil e, às vezes, impossível. A classificação de uma resposta no nível de conhecimento ou raciocínio depende de como o aluno foi ensinado e não, apenas, da questão apresentada. Aplicar um teste somente para verificar se o aluno “sabe” ou “não sabe” um certo assunto não é inteiramente relevante para fins de avaliação de um curso, conforme Cronbach, importando, isto sim, medir o conhecimento em termos de profundidade, relacionamento e capacidade de aplicá-lo a novas situações.

A questão da especificidade das questões é considerada por Cronbach, que discute detalhadamente o problema. A especificidade, na maior parte das vezes, concentra-se no uso de uma terminologia própria do curso, que somente é compreendida por aqueles que tiveram a oportunidade de assistir a ele. Ainda que o conhecimento dessa terminologia seja importante, é mais importante, para fins de avaliação, a medida de compreensão de relação e de outras variáveis do curso, da que, em princípio, poderia ser verificada, também, em quem não foi diretamente submetido ao curso em questão.

O trabalho de Cronbach permite, entre outras, as seguintes conclusões:

1. A avaliação educacional requer a descrição de resultados; dessa forma, certas preocupações das medidas educacionais para a produção de escores precisos, visando a comparar indivíduos ou escores médios de diferentes cursos, pouco contribui para a avaliação educacional, cuja descrição dos resultados deve ser a mais ampla possível, ainda que às custas do sacrifício de uma suposta justiça e precisão.
2. A análise do desempenho em itens isolados ou em certos tipos de problemas fornece mais informações do que a análise de escores compósitos.
3. O objetivo da avaliação educacional não consiste em, simplesmente, aquilatar o valor de cursos, rejeitando-os ou aceitando-os, mas, sim, em ser uma parte fundamental no processo de desenvolvimento de currículos, através da coleta e do uso de dados que possibilitem uma compreensão mais profunda do processo educacional (VIANNA, 1982b).

A avaliação necessita considerar aspectos do meio educacional em que se desenvolverá, evitando, assim, a imposição de modelos nem sempre ajustáveis aos vários aspectos do sistema, ainda que bem estruturados do ponto de vista teórico. A realidade brasileira possui aspectos de grande singularidade, sendo preferível a adoção de algumas ideias gerais para orientação e fundamentação do trabalho de avaliação, como as que foram desenvolvidas por Cronbach (1963). Essas ideias são perfeitamente válidas em função de nosso contexto, mas não constituem, exatamente, um modelo. A adoção de um modelo transplantado de outras culturas pode criar uma situação artificial, sem desdobramentos práticos.

É preciso, no processo de desenvolvimento de uma avaliação, compreender o sentido de certos conceitos psicométricos, como validade, fidedignidade e discriminação, importantes sem dúvida em medidas educacionais, os quais, entretanto, devem ser ponderados em função de uma avaliação que não vise à seleção e ao prognóstico de comportamentos.

A validade de conteúdo é geralmente associada a uma amostragem representativa de conhecimentos e comportamentos adquiridos durante o processo educacional. A conceituação não é

pacífica, havendo quem exclua os comportamentos, limitando o problema à representatividade dos conteúdos programáticos. A questão dos comportamentos (objetivos instrucionais/educacionais) é, por sua vez, extremamente controversa. Há, inclusive, quem reaja ao ensino e à avaliação por objetivos. Sem entrar no mérito da questão, mas dentro da linha de considerar, na validade curricular, a simultaneidade dos conteúdos e dos objetivos, é preciso atentar para algumas particularidades desses últimos e para suas implicações no processo de avaliação. O fato de diversas taxonomias assinalarem níveis vários de objetivos não significa que, necessariamente, todos esses objetivos devam ser avaliados. Apenas aqueles que foram desenvolvidos no decorrer do processo instrucional devem ser, efetivamente, considerados. O desejável, às vezes, não corresponde à realidade, e um dos objetivos da avaliação é, justamente, descrever essa realidade tal como se revela por intermédio dos dados coletados. Ainda, com referência a comportamentos (objetivos), impõe-se atentar para o fato de que os mesmos são hierarquizados em ordem crescente de complexidade e abstração, sendo impossível, com bastante frequência, obter a concordância de especialistas sobre comportamentos mais heterogêneos, como, por exemplo, o de “avaliar” ou de “analisar”. Desse modo, é comum, em avaliação educacional, a concentração de todos os comportamentos gerais e abstratos em uma única categoria: aplicação.

O fato de medir com fidedignidade não significa que se esteja realizando uma boa avaliação. A precisão com que se mede é condição necessária, mas não suficiente para a excelência de uma mensuração. Entre os atributos dos instrumentos de medida, importa realmente a validade, estando aí implícita a fidedignidade, não sendo verdadeira, entretanto, a situação recíproca. A fidedignidade, qualquer que seja a técnica usada para estabelecê-la (Alpha de Cronbach ou seu caso particular, as fórmulas de KuderRichardson), depende, fundamentalmente, da variância dos resultados dos escores. Quanto maior a variância, em princípio, maior será a precisão das medidas. Se esse aspecto é capital em uma avaliação somativa ou em um processo de seleção que, no fundo, é também uma avaliação de produto final, o mesmo não ocorre na avaliação formativa e/ou em uma investigação voltada para a avaliação de programas, a partir do desempenho escolar. O importante é a homogeneidade do grupo avaliado, que

refletiria domínio do material instrucional, entre outros aspectos, e não a heterogeneidade resultante da maior variabilidade entre os vários níveis de desempenho.

O cálculo de coeficientes de discriminação, expressos por diferentes tipos de medidas de correlação (*r*-bisserial, *phi*, entre outros), é presença obrigatória em estudos psicométricos na área da docimologia educacional, quando, especialmente, há necessidade de contrastar desempenhos extremos, a fim de verificar o comportamento dos itens/questões do teste/prova na separação dos indivíduos com vários níveis de domínio instrucional. A teoria clássica das medidas põe em destaque a importância desses coeficientes, favorecendo os itens/questões que discriminaram positivamente, ou seja, aqueles itens/questões que foram respondidos, corretamente, pela maioria dos que obtiveram escores elevados e por poucos ou nenhum dos elementos situados no extremo inferior da distribuição de escores, em oposição aos itens/questões de discriminação negativa, isto é, que foram respondidos, em maior proporção, pelo grupo de elementos situados no extremo inferior da distribuição de escores e por pequena proporção dos situados no extremo superior. Os primeiros são valorizados por terem cumprido o seu papel discriminar; os segundos, condenados, pois não concretizaram o esperado discriminar. A avaliação formativa, assim como a avaliação com vistas a um determinado programa ou currículo, não objetiva discriminar, não pretende estabelecer diferentes níveis de desempenho. Ao contrário, a não discriminação, a homogeneidade do grupo, o máximo de respostas corretas, respeitadas as diferenças individuais, é o objetivo primordial de um ensino eficiente e de um programa ou currículo estruturado adequadamente, porquanto demonstraria sensibilidade ao processo instrucional e à sequência curricular. Ainda que possam constar de estudos de avaliação educacional, nem sempre os coeficientes de discriminação são peça essencial, apesar de necessários para análise de questões em uma avaliação somativa.

Um problema, muitas vezes, apresenta-se em avaliação educacional: é necessária a construção de testes padronizados e o uso de diferentes tipos de normas em investigações sobre o rendimento escolar? As palavras “testes padronizados” e “normas”, quase sempre, são usadas com imprecisão, con-

fundindo aplicação dos instrumentos de medida em condições padronizadas e segundo determinadas normas (orientações) com um teste planejado para proporcionar uma amostra sistemática do desempenho individual, aplicado segundo determinadas instruções, corrigido em conformidade com certas regras e interpretado por meio de informações normativas. A norma, por sua vez, é um conjunto de valores típicos descritivos do desempenho, num determinado teste, de um grupo específico de indivíduos, supostamente representativos de uma certa população. O contexto brasileiro, felizmente, não conhece os testes padronizados na área da educação. As restritas experiências realizadas no passado não prosperaram. Voltando à pergunta inicial, a resposta é negativa. A avaliação educacional pode, perfeitamente, prescindir de testes padronizados, que, além de extremamente caros, exigindo investimento de somas vultosas na sua construção, são demasiadamente genéricos e não possibilitam verificar o desenvolvimento programático realizado em sala pelo professor, no seu dia-adia. Os melhores instrumentos são as provas *ad hoc*, elaboradas por professores, com base na sua experiência, e revistas por outros professores, a partir de suas vivências pessoais. A avaliação educacional também não visa a desenvolver testes que sirvam *urbe et orbi*. Ainda que, teoricamente, isso seja possível, não faria sentido tal esforço e investimento, quando outros objetivos são prioritários em educação e a avaliação educacional pode contribuir para a sua concretização.

A avaliação educacional, segundo a conceituação apresentada, que a situou como um conjunto de operações ‘visando levantar informações que possibilitem uma tomada de decisões, como usar determinado livro, modificar um currículo, identificar deficiências no processo de aprendizagem, positivar falta de sensibilidade à instrução, pode adotar como amostra diferentes tipos de conjuntos: um pequeno grupo de alunos, uma série completa, alunos de um ou vários tipos de colégios, estudantes de diferentes áreas geográficas etc. A amostragem na avaliação educacional, assim como em outros tipos de investigação na área da educação, sofre limitações decorrentes do tempo, de pessoal e, especialmente, de orçamento. Todos esses fatores devem ser considerados. A questão tempo é, muitas vezes, a mais crucial, pois a tomada de decisões exige uma

coleta rápida para ação imediata, a fim de cortar a propagação de maiores danos em decorrência da demora da intervenção no processo educativo. A problemática de pessoal precisa ser considerada quando a natureza do trabalho exige material humano especializado, essencialmente na fase de aplicação dos instrumentos. A avaliação não é uma área de concentração, e poucos são os que, efetivamente, a ela se dedicam. A situação atinge contornos dramáticos quando os trabalhos de avaliação envolvem grandes regiões geográficas, às vezes de difícil acesso. A tudo isso acrescentam-se os custos da avaliação, que a tornam inviável em termos orçamentários, apesar de perfeitamente desenhada segundo as técnicas de amostragem. Assim, muitas vezes, seria desejável que todas as áreas geográficas estivessem adequadamente representadas com seus diferentes tipos de escolas, formando assim estratos nos quais, posteriormente, de forma aleatória, seriam selecionadas as instituições e indivíduos que, finalmente, integrariam a amostra. Nem sempre esse procedimento – que, supostamente, garantiria a representatividade da amostra – possibilita a factibilidade da avaliação em termos operacionais. Ao avaliador cabe decidir entre o ideal, fundamentado na matemática da amostragem mas irrealizável em termos concretos, e o possível em termos de realidade, que pode levar a inferências, a ilações, a implicações e, a partir dos dados levantados, chegar a conclusões que oportunizem decisões e modificações da realidade, objetivo final do processo de avaliação.

A adoção de um modelo matemático de amostragem seria necessária se a avaliação educacional, com o objetivo de identificar deficiências no rendimento escolar, por exemplo, visasse à generalização dos resultados a todo o sistema e não restringisse suas conclusões ao grupo de escolas e alunos participantes da amostra. Importante, no caso específico de uma avaliação educacional com vistas ao levantamento de “competências cognitivas” dos estudantes, é considerar a situação crítica das estatísticas sobre educação no Brasil. A estruturação de uma amostra para fins de avaliação educacional deveria “ter um conhecimento bastante seguro dos índices de repetência, promoção e evasão existentes no país”, conforme Fletcher e Ribeiro (1988); contudo, é preciso convir que as discrepâncias entre as estatísticas educacionais coletadas por diferentes órgãos governamentais, em decorrência da diversidade das metodologias empregadas, são consideráveis,

chegando muitas vezes a 40%, como é o caso das estatísticas sobre população em idade escolar fora do sistema educacional. A questão da amostragem em avaliação educacional é complexa e, às vezes, difícil de solucionar, porque, malgrado o rigor dos procedimentos lógico-estatísticos, sempre subsiste a pergunta: a amostra selecionada é representativa do universo investigado?

O importante, em relação à atual estrutura de ensino em seus vários graus, é criar um sistema de avaliação externa que, de modo sistemático, informe aos responsáveis pelo ensino/educação os problemas da realidade pedagógica que ocorrem na escola e que se refletem nos diferentes níveis de capacitação cognitiva. A identificação de pontos críticos no desempenho escolar dos estudantes deve ser um dos objetivos do diagnóstico de deficiências, para que possam ser corrigidos os desvirtuamentos do processo ensino/aprendizagem.

O praticante da avaliação educacional, frequentemente, necessita posicionar-se em relação a problemas aparentemente menores mas que perturbam o seu trabalho, se não forem equacionados. As decisões sobre esses problemas são subjetivas na maioria das vezes. A vivência do avaliador e seu bom senso encontrarão uma solução para as situações que, imprevisivelmente, possam surgir. Assim, o avaliador, na construção dos instrumentos de mensuração, vê-se às voltas com a definição de um conteúdo programático que corresponda à média do que ocorre nas escolas do sistema, no caso de uma avaliação do rendimento. As Secretarias de Estado da Educação poderiam colaborar na solução do problema, mas, quase sempre, acham-se impossibilitadas, por não possuírem cópias dos programas desenvolvidos nas escolas. A legislação permite que cada escola estabeleça seus próprios conteúdos, havendo, portanto, grande diversidade de sequências curriculares em escolas de uma mesma cidade situadas próximas uma das outras. Algumas Secretarias possuem guias curriculares que oferecem uma certa ordenação ao conteúdo em suas linhas gerais, mas essa situação não é a regra comum. O avaliador vê-se compelido a criar um programa mínimo, não muito coerente às vezes, a partir dos aspectos comuns às várias programações. Uma solução alternativa e parcial para o problema estaria na análise dos livros didáticos utilizados, que se constituem em fonte subsidiária de informações sobre o que é desenvolvido em sala de aula pelos professores. O procedimento não

foge à prática ortodoxa da avaliação, sendo usado, inclusive, em outros contextos educacionais, para a construção dos chamados testes padronizados. Ajusta-se perfeitamente à nossa realidade, pois é sabido que, em muitas escolas, o livro didático é o verdadeiro “programa” curricular, como já foi comprovado em pesquisas nacionais. O livro didático, ainda que elemento colateral de informação, é um instrumento importante para que o avaliador possa caracterizar o que, efetivamente, ocorre no mundo da escola e, especialmente, em sala de aula.

A participação de professores na avaliação educacional, por intermédio da elaboração de questões e de objetivos instrucionais para uso futuro em investigações, constitui motivo de preocupação para alguns especialistas. O problema precisa ser dimensionado levando em consideração o tipo de avaliação e seus objetivos. Se a avaliação é interna, com finalidade formativa ou somativa, ninguém melhor do que o próprio professor para avaliar seus alunos, tendo em vista que foi ele quem estabeleceu objetivos, usou estratégias para conseguir o máximo de aprendizagem e orientou o aluno ao longo do processo instrucional. Se a avaliação é externa, usar professores da própria escola envolvida na investigação seria criar um novo problema, que desvirtuaria os resultados, levantando elementos que comprometeriam a tomada de decisão. As medidas realizadas seriam, sem sombra de dúvida, fidedignas, em grau talvez bastante elevado, mas a validade estaria irremediavelmente prejudicada, a não ser que se pudesse organizar um *pool* de professores igualmente capacitados, que gerariam um *pool* de questões de boa qualidade. Posteriormente, esses itens seriam selecionados de forma a constituir uma amostra representativa embasada na relação professores/alunos/escolas/currículos/objetivos/metodologias que integram a avaliação externa. A realidade educacional enfrentada pelos avaliadores é bem diversa do ideal, exigindo, assim, a adoção de práticas acauteladoras para o bom êxito do trabalho. A utilização de professores integrados no sistema educacional, independentemente da região ou área geográfica em que exerçam suas atividades, não contribui para desfigurar o processo de avaliação, desde que sua experiência resulte da prática objetiva do ensino, na realidade do dia a dia.

O problema do acesso às provas que serão aplicadas nas avaliações externas é conexo ao da elaboração de questões/obje-

tivos pelos professores das instituições avaliadas. Algumas experiências em avaliação educacional perderam o seu significado, justamente em face desse acesso. Ao tomar conhecimento das áreas de conteúdo objeto da avaliação, o professor, inconscientemente, tende a acentuar esses aspectos, provocando, portanto, um direcionamento favorável a seus alunos, em oposição aos demais sujeitos submetidos à avaliação. A avaliação externa, se não for bem compreendida pelos professores, gera uma certa ansiedade entre eles, em decorrência, possivelmente, do atual caráter punitivo de que, muitas vezes, reveste-se a avaliação em nosso contexto. O professor, desse modo, tende a criar uma imagem favorável para seu desempenho através do bom desempenho dos seus alunos. A avaliação procura obter dados para orientar o professor em seu trabalho, indicando as distorções no processo de aprendizagem. Busca, também, outros elementos que facilitem o processo decisório, não vindo a criar uma situação de constrangimento para o professor, a escola e o próprio sistema educacional.

A divulgação das provas usadas em pesquisas ligadas diretamente à avaliação costuma ser advogada por alguns elementos da comunidade educacional, criando uma situação de constrangimento para o avaliador que, a duras penas, construiu seu instrumental. Há uma tradição em nosso contexto social favorável à divulgação de provas e testes, inclusive por intermédio da imprensa, o que precisaria ser rompido. Ao avaliador cabe decidir o momento apropriado para divulgação do instrumental, no todo ou em parte. Alguns relatórios reproduzem fragmentos com tipos específicos de questões, outros se limitam à descrição geral do instrumental e, ainda, um terceiro grupo o divulga na sua inteireza, impossibilitando, desse modo, a replicação dos mesmos instrumentos em sucessivas avaliações. Ao longo do tempo, com a ampla divulgação dessas provas e a grande rigidez curricular, torna-se difícil a construção de novos instrumentos de medida sem o comprometimento da sua validade. Por outro lado, a disseminação de provas favorece o treinamento para a avaliação, como já vem ocorrendo em outros contextos que, tradicionalmente, usam testes padronizados. Ao avaliador cabe julgar, portanto, após considerar diferentes variáveis, o acesso às provas e a divulgação das mesmas. A prudência recomenda, entretanto, que o sigilo seja mantido,

se o objetivo for criar uma tradição relativa à avaliação externa, tornando-a parte integrante da vida escolar.

O uso de provas objetivas ou discursivas costuma ser apresentado como um problema para alguns avaliadores. A questão, em si, é um falso dilema, considerando-se que existe extensa literatura com base empírica, demonstrando, de modo insofismável, a adequação de ambos os tipos de prova para medir praticamente os mesmos traços, sendo, pois, inteiramente ociosa qualquer discussão sobre o assunto. Ao avaliador o problema deve se apresentar de uma outra forma, considerando a natureza da investigação e a população avaliada. Dependendo do contexto, talvez sejam recomendados instrumentos semiobjetivos apenas ou, então, inteiramente objetivos, em função do tempo disponível, da problemática relativa à sua aplicação e aos custos operacionais. A discussão sobre itens discursivos/objetivos não é bizantina, como poderia parecer à primeira vista, merece reflexão, mas não deve ser colocada em termos extremados, pois depende de um concurso de variáveis que interagem, cabendo assim ao avaliador tomar a decisão que melhor se ajuste ao quadro investigativo.

A avaliação, segundo a perspectiva ora desenvolvida, é um processo sistemático de levantamento de dados relativos a um determinado fenômeno, a fim de possibilitar a tomada de decisões com base em julgamentos de valor. A avaliação educacional aplicada à situação específica do rendimento escolar procura, desse modo, identificar as necessidades instrucionais dos alunos, com base em pontos críticos do seu desempenho em provas de escolaridade. Busca coletar elementos para fundamentar a análise da eficiência do sistema de ensino. A avaliação educacional está, basicamente, comprometida com a melhoria da qualidade do ensino e com a compreensão da influência dos professores no desenvolvimento de programas educacionais.

REFERÊNCIAS BIBLIOGRÁFICAS

BLOOM, B. S. et al. *Taxonomy of educational objectives: handbook 1; Cognitive domain*. New York: David McKay, 1956.

CRONBACH, L. J. Course improvement through evaluation. *Teachers College Record*, New York, v. 64, n. 8, p. 672-683, May 1963.

FLETCHER, P. R.; RIBEIRO, S. C. *A educação na estatística nacional*. In: SEMINÁRIO DE AVALIAÇÃO DAS PNADS, 1980. Nova Friburgo: ABEP, 1988. Mimeo.

HAMMOND, D. L. *Evaluation at the local level*. Tuckson, Arizona: EPIC Evaluation Center, 1967.

METFESSEL, N. S.; MICHAEL, W. B. A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. *Educational and Psychological Measurement*, v. 27, 1967.

POPHAM, W. J. *Educational evaluation*. Englewood Cliffs (New Jersey): Prentice Hall, 1975.

PROVUS, M. M. *Discrepancy evaluation*. Berkeley: McCutcham, 1971.

SCRIVEN, Michael. The methodology of evaluation. In: STAKE, R. E. (ed.) *Curriculum evaluation*. Chicago: Rand McNally, 1967. (AERA. Monograph, n.1).

STAKE, Robert E. The countenance of educational evaluation. *Teachers College Record*, v. 68, n. 7, p. 523-540. 1967

STUFFLEBEAM, D. H. et al. *Educational evaluation and decision making*. Itasca, Illinois, Peacock, 1971.

TYLER, R. W. General statement on evaluation. *Journal of Educational Research*, v. 35, p. 492-501, 1942.

VIANNA, Heraldo M. Avaliação educacional: algumas ideias precursoras. *Educação e Seleção*, São Paulo, n. 6, p. 63-70, 1982b.

_____. Avaliação educacional: problemas gerais e formação do avaliador. *Educação e Seleção*, São Paulo, n. 5, p. 9-14, 1982a.

WORTHEN, B. R.; SANDERS, J. R. *Educational evaluation: theory and practice*. Washigton, DC: Charles A. Jones, 1973. 372 p.