

TESTES E AVALIAÇÕES DO ENSINO NO BRASIL

Bernardete A. Gatti*

Avaliação é um conceito que apresenta uma conotação emocional muito grande na cultura brasileira. Não se fala impunemente em avaliar. Cria-se um clima de insegurança, defesa, até de agressividade e hostilidade. Falar-se em avaliação é criar uma situação de ameaça. ⁽¹⁾Questão de cultura? Questão de associação com situações aversivas? Questão ligada a posições políticas? Muito haveria para se explorar neste segmento. No entanto, o que interessa à presente reflexão é reter este dado que interfere até mesmo nas discussões acadêmicas. As polêmicas sobre os exames vestibulares e as avaliações no ensino em geral ilustram bem este ponto. Toda avaliação envolve, é certo, um relativo grau de ansiedade. Porém, o mal estar gerado em nosso meio por essas situações assume proporções que vão além do esperado. Por isso, o exame desta questão sofre interferências emocionais não desprezíveis.

As situações em que a avaliação se apresenta como necessária, enquanto processo de levantamento de evidências confiáveis, sob certa perspectiva, são múltiplas. Ao falarmos em avaliação do ensino precisamos situar em que nível de preocupação colocamos a questão: se a nível de sistema ou sub-sistemas, se a nível de programas educacionais; se a nível de pesquisa educacional ou se a nível de sala de aula. Precisamos ainda ter claro a que se destina essa avaliação, quais seus objetivos. Confor-

* Do Departamento de Pesquisas Educacionais da Fundação Carlos Chagas

⁽¹⁾Esta questão é examinada, por exemplo, por Maria Amélia A. Goldberg em dois artigos publicados no nº 1, julho de 1980, da revista *Educação e Avaliação* (Cortez Editora, São Paulo), págs. 23-47; págs. 96-117.

me os objetivos e o nível de abrangência, sem dúvida, a conceituação, os tipos, as funções, as técnicas e os instrumentos de avaliação são diferenciados. Além de diferenciadas as possíveis abordagens teóricas com que a avaliação mesma pode ser enfocada.

Os testes em avaliação de ensino constituem uma faceta particular da questão da avaliação educacional. A experiência brasileira com avaliação educacional, de modo geral, é bem mais ampla e diversificada — tanto no uso como na quantidade de estudos e pesquisas — do que a experiência com testes educacionais.

Examinando 312 trabalhos produzidos no Brasil sobre avaliação educacional entre 1960-1986, compreendendo livros, artigos de revistas, dissertações de mestrado e teses de doutoramento verificamos que apenas 7% tratam de algum aspecto que envolve a construção ou a pesquisa sobre testes educacionais. A pesquisa sobre testes, sua construção, uso e impacto é sobremaneira rarefeita entre nós. Ou seja, o conhecimento científico que temos nessa área é pequeno, donde concluímos que os debates que se fazem em nosso meio, em torno de questão tão séria e de tal importância social, têm pouca fundamentação científica. Sua base parece ser mais impressionística.

O uso de testes educacionais de modo mais generalizado se deu no Brasil a partir de meados da década de 60. Seu emprego em nosso meio está muito associado a exames vestibulares, através dos quais esta maneira de avaliar conseguiu ampla divulgação. É, também, neste período que uma perspectiva mais tecnicista e economicista começa a dominar na área educacional, principiando a proliferação de textos mais específicos sobre medidas educacionais, nos quais se discutem as técnicas de elaboração das chamadas questões e provas objetivas e as questões estatísticas envolvidas na teoria das medidas.

No entanto, assistimos a um paradoxo. Se de um lado empregam-se os testes para avaliar os candidatos ao ensino superior em todo o país, o impacto dentro das escolas superiores de cursos de medidas educacionais foi pequeno, sendo incluído na maioria dos casos como um único curso de duração de um semestre. Ou seja, não se propiciou uma formação adequada de recursos humanos na área. A preparação destes recursos e o desenvolvimento de competência no que se refere a medidas educacionais enfrentou, além disso, a tão conhecida aversão dos estudantes e professores das áreas das ciências humanas por todo conhecimento que se aproxime do quantitativo ou que empregue raciocínios de base matemática, por mais simples que sejam. Se alguns especialistas foram formados, sobretudo recorrendo-se a cursos no exterior, estes se constituíram em pequeno núcleo de competência, competência esta que ficou esparsa entre algumas instituições, em meio a um grande número de profissionais pouco preparados e com pouca base para uma reflexão sobre os fundamen-

tos lógicos e metodológicos desta técnica de avaliação. Isto levou, por exemplo, à identificação, no senso comum, de prova objetiva com “teste de cruzinha”, quando se sabe que a maioria das formas que podem assumir as questões objetivas não são as de escolha múltipla e que, dominados os fundamentos da técnica, a criação de novas formas é uma questão de exercício dessa competência no trato de conteúdos específicos. Em qualquer das formas que uma questão objetiva possa assumir, é possível verificar a capacidade de raciocínio, de análise, de pensamento crítico e de outras características complexas de que se reveste a aprendizagem humana. Mas para produzir este tipo de teste uma formação adequada é indispensável. Uma visão crítica também, ao lado de um razoável conhecimento da realidade educacional e, como coloca Vianna (1986), implica “uma filosofia educacional centrada na pessoa do próprio estudante”. (pág. 99). É ainda Vianna que diz: “as provas objetivas utilizadas nos vestibulares a partir de 1965 baseavam-se no pressuposto, comprovado empiricamente, de que poderiam medir comportamentos complexos, se fossem bem elaboradas”, e, desse modo, traçar um perfil do desempenho escolar a nível de 2.º grau” (pág. 102). Acrescentando que, nesse período, “assim como se nota um acentuado experimentalismo em nosso contexto educacional, ..., observa-se, igualmente, acentuada tendência à improvisação com vistas ao atendimento de certos modismos. A partir da década de 60 os vestibulares começaram a utilizar de modo indiscriminado questões objetivas, sob a forma de múltipla escolha, sem que os envolvidos no processo de elaboração, na maioria das vezes, tivessem suficiente *expertise* para o desenvolvimento de um trabalho plenamente satisfatório.” (pág. 102).

Por outro lado, a nível dos livros didáticos destinados a alunos de 1.º e 2.º graus, começam a aparecer nos textos, em profusão, testes de escolha múltipla propostos a título de exercício e verificação, quando não substituindo o próprio texto. Observa-se que isso ocorreu praticamente em todas as áreas de conteúdos específicos (língua portuguesa, estudos sociais, matemática etc.), com acentuada presença nos livros didáticos destinados às quatro primeiras séries do 1.º grau. O exame da estrutura desses testes deixa claro o despreparo dos autores no que diz respeito à construção deste tipo de medida para avaliação do ensino. Sua elaboração é calcada no senso comum, uma tradução barata de questões objetivas de escolha múltipla, oferecendo lamentável imagem de uma prova objetiva, com itens envolvendo sobretudo a memorização imediata, sem propor o acionamento das funções mais complexas de análise e síntese, por exemplo, e outras dessa natureza, estas sim, fundamentais no desen-

(*) Grifo nosso.

volvimento e, portanto, na avaliação do ensino ou de funções cognitivas importantes na aprendizagem.

Com o recrudescimento das críticas a esse uso indiscriminado dos testes de múltipla escolha, generaliza-se sua condenação sem um exame mais acurado e científico de suas raízes, e sem desvelar as reais distorções envolvidas na sua produção e generalização. Não se examinou se os efeitos apontados como nocivos à aprendizagem (redução à memorização, desuso da escrita e da expressividade lingüística, entre outros) tinham suas causas *nestas* formas empobrecidas e mal elaboradas de testes objetivos. Nem se fez o estudo de quais os efeitos no ensino, ou na aprendizagem, quando se usam questões objetivas construídas dentro de pressupostos educacionais claros e com o rigor necessário às premissas de seu emprego, em suas diferentes modalidades. Nem se questionou se o que estava sendo feito e usado neste tipo de avaliação era, ou é, conceitualmente, uma prova objetiva ou o quê. A qualidade dos testes não foi questionada nem estudada. A condenação foi em tese. Nisto interferiu a quase total ausência de especialistas na área, o tônus emocional de que se reveste para a classe média brasileira a seleção para o ensino superior, e, sem dúvida, algumas questões de ordem política geral e disputa de poderes locais. Nesse sentido, o alvo maior das disputas foram sempre os testes dos vestibulares, pouco se falando, e nada se analisando, sobre os livros didáticos. Com essa tônica generalizou-se apenas a crítica, não se discriminando a que aspectos ela se dirigia e sobre que tipo de medida realmente se reportava. Aliás, apenas como efeito ilustrativo, mas que traz um ponto de reflexão crítico e que não tem sido encarado de frente, quer pelos normatizadores e administradores, quer pelos interessados, é o impacto nulo da introdução da redação nos exames vestibulares na melhoria do desempenho em língua portuguesa dos candidatos aos cursos superiores. Tornada obrigatória a partir de 1978, com a perspectiva de que influísse para melhor desempenho na expressão escrita dos estudantes, não causou de modo algum o impacto esperado, e provocou a introdução, na avaliação dos candidatos, de um alto índice de arbitrariedade e variabilidade nos julgamentos, uma vez que intervem nessa avaliação um grande número de avaliadores cuja parametrização de critério é praticamente impossível de garantir. [Teixeira (1980); Ribeiro e outros (1981); Vianna (1981); Sanchez e outros (1981); Rodrigues e outros (1982); Mello (1982); Ripper (1983); Buchweitz e outros (1984); Vianna (1984).] A introdução nesse processo seletivo de questões discursivas caminha na mesma direção. Não nos cabe aqui discutir a relevância ou não desta introdução, mas lembrar que os aspectos qualitativos envolvidos na elaboração de uma questão objetiva são os mesmos para a elaboração de uma questão discursiva do ponto de vista da avaliação educacional. E, nesta nova proposta de medida do ensino e da aprendizagem, a qualidade das

questões discursivas precisa ser analisada e discutida, porque, como coloca Vianna (1986), “se antes havia o problema da questão objetiva elaborada de maneira deficiente, agora existem dois problemas, pois ao anterior se agrega o da elaboração igualmente insatisfatória das questões discursivas. Além disso, surgem indagações sobre o subjetivismo das correções, que não tem maior relevância no contexto do dia-a-dia da sala de aula, mas exerce influência desastrosa em exames de massa como o vestibular” (pág. 103).

Sob uma certa ótica os exames vestibulares se constituem em uma avaliação do ensino nos níveis anteriores e caberia uma reflexão mais acurada sobre isto. Da perspectiva de avaliações diretas do ensino a nível de 1º e 2º graus, sob a ótica do rendimento escolar, compreendendo um sistema ou sub-sistema visando a subsidiar decisões de administração educacional, de políticas de desenvolvimento de recursos humanos, de estruturas curriculares e outras, há uma grande lacuna entre nós. Poucos se aventuraram na área e precisamos reconhecer que a questão não é nada simples. Mas devemos reconhecer também que os poderes instituídos não se têm interessado por isso e nem os componentes das redes escolares têm tranqüilidade, ou pelo menos boa disposição, para encarar esse tipo de avaliação. Por mais que os objetivos colocados sejam de ajuda, de reconhecimento de terreno e de oferta de dados palpáveis para o planejamento do ensino, a interpretação é sempre uma só: a de controle e punição pela exposição de “fraquezas”. Nem por isto deixou-se de fazer algumas tentativas.

No final da década de 60 e princípios de 70, o Centro de Estudos de Testes e Pesquisas Psicométricas, da Fundação Getúlio Vargas, no Rio de Janeiro, empreendeu a construção de um conjunto de testes de desenvolvimento educacional (TDE), com o objetivo de avaliar o rendimento escolar de alunos da escola de nível médio (2º grau), no então Estado da Guanabara. Elaborou-se e padronizou-se um conjunto de testes de linguagem, matemática, ciências físicas e naturais e estudos sociais. Esperava-se que estes testes pudessem ser utilizados para exercer uma influência renovadora na relação de ensino em sala-de-aula. Construiu-se um manual e seu uso deveria ser feito por técnicos de educação, que poderiam utilizar os testes para auxiliar a orientação de alunos e de professores. A construção da bateria TDE, na expressão dos responsáveis pela sua elaboração, “não teve um fim em si mesma — muito ao contrário, o processo de elaboração dos testes integrou-se num programa de seminários, de cursos, de bolsas-de-estudo e publicações, que procurava o aperfeiçoamento do professor. Através de seminários, principalmente, desenvolveu-se a discussão dos objetivos da escola média e a tentativa de traduzí-las por definições em termos de questões de testes” (FGV-CETPP, 1971, pág. 3). O estudo cuidadoso e amplo que foi feito para a elabora-

ção dos itens, sua seleção, aplicação para padronização, validação e verificação da fidedignidade foi acompanhado de uma pesquisa sobre os efeitos de diferentes variáveis sociais e econômicas em seus resultados. Examinando os dados desses estudos tem-se um panorama claro, dentro de determinados objetivos educacionais, do rendimento escolar dos alunos de grau médio no Estado da Guanabara, neste período, por curso, turno e zona de escola, bem como por outras variáveis (instrução do pai, renda familiar etc.). Além disto, padronizaram-se instrumentos que poderiam ter utilização futura. Que utilização destes dados fez a administração da educação? Nenhuma. Enquanto o estudo se processou houve envolvimento de técnicos e professores das áreas e colégios em questão e um impacto localizado sem dúvida aconteceu. E depois? Qual a ampliação de seu emprego, revisões, atualizações, críticas?

Em 1980, o Departamento de Planejamento e Orientação da Secretaria Municipal de Ensino, da Prefeitura do Município de São Paulo, realizou pesquisa para verificação do nível de escolaridade dos alunos da rede municipal de ensino. Foram estudados o 3º estágio de Educação Infantil, e as 1ªs, 3ªs, 5ªs e 7ªs séries do 1º grau, por intermédio de questões objetivas (não necessariamente de múltipla escolha) de língua portuguesa, matemática e ciências, construídas e testadas para esta finalidade. O estudo tinha o objetivo de “subsidiar os órgãos técnicos no sentido de fornecer material de apoio para a revisão e implementação das propostas curriculares das Escolas Municipais de 1º grau e de Educação Infantil” e “oferecer dados e informações aos órgãos técnico-administrativos responsáveis por treinamento, orientação e supervisão de Educação Infantil de 1º grau” (SME-DEPLAN, 1981). Os dados foram analisados no todo e por Delegacia de Ensino Municipal (DREM), sendo discutidos em seminários especiais com os técnicos dos diferentes órgãos centrais e com cada DREM. Os debates foram muito vivos, não sem as atitudes defensivas costumeiras, porque mesmo que a colocação seja: “sob este ângulo, o que temos são estes dados; vamos usá-los para uma reflexão e projeção de ação”. A tradução é: “Veja como está ruim o rendimento escolar em sua delegacia de ensino”, com a conseqüente culpabilização. No entanto, com estas reuniões alguns passos foram dados no sentido de se trabalhar melhor este tipo de avaliação. Algum impacto este trabalho teve nesse momento, em que algumas propostas de materiais e métodos de ensino estavam sendo revistas. E então? Muda a administração, mudam-se as equipes, não se recupera o caminho andado nem mesmo para análise, revisão ou crítica, e os dados são perdidos. A experiência dos técnicos e professores também.

Uma avaliação de rendimento escolar de porte mais amplo, através de questões objetivas foi feita como parte do projeto de avaliação de implementação do programa *EDURURAL*, no Nordeste brasileiro, pela

Fundação Cearense de Pesquisa e Cultura (FCPC, 1982, 1984, 1987), em convênio com a Fundação Carlos Chagas. Destinou-se a aquilatar a aprendizagem escolar das crianças até a 4.^a série do 1.^o grau, em língua portuguesa e matemática e sua alteração ou não em função dos investimentos na área de treinamento de professores, material didático, construções escolares etc., ao longo de cinco anos (80-85). Foram colhidos dados, em sistema de amostragem, em três Estados: Piauí, Ceará e Pernambuco, nos anos de 1981, 1983 e 1985, com formas paralelas das provas. Os resultados de 1981 e 1983 foram discutidos com os responsáveis pelo projeto no MEC e com as respectivas Secretarias nos Estados, sendo-lhes fornecidos relatórios detalhados. Os dados de 1985 estão analisados, mas ainda não foram discutidos. O objetivo do acompanhamento ao longo dos anos era oferecer às equipes do *EDURURAL* um material mais objetivo para reflexão, para crítica, orientação e redirecionamento de ações. Pouco uso se fez até aqui desses dados, e os pontos críticos apontados na seqüência dos relatórios, no que dizem respeito a aspectos de ensino de língua portuguesa (alfabetização) e matemática, que oferecem ganchos interessantes para uma atuação mais eficaz na área, ou ainda para uma reconsideração da questão e das ações, parecem não ter recebido dos técnicos a atenção necessária e um estudo mais detalhado. Saberão eles "ler" estes dados? Ir além da primeira informação? Que condições teriam de passar dessa compreensão para a criação de alternativas de soluções e de ações concretas? Ou o problema se situa num plano político? Que autonomia têm para atuar no Projeto?

A interpretação, análise e crítica de um processo avaliativo requer algumas condições que passam não só pelo conhecimento e domínio de técnicas de medidas educacionais e modelos de avaliação, como pelo conhecimento dos conteúdos envolvidos, da realidade a que se reportam, dos pressupostos educacionais em que se baseiam e dos objetivos de ensino privilegiados.

O professor em sua sala de aula vive tomando decisões sobre o como e o quê ensinar a cada momento, em que seqüência, de que modo, e está de um modo ou de outro constantemente avaliando suas ações e os alunos. Os técnicos educacionais tomam amiúde decisões sobre como orientar professores e alunos, que subsídios oferecer, que técnicas privilegiar. O administrador e o planejador, a nível de sistemas e sub-sistemas, estão constantemente assumindo decisões sobre políticas educacionais, propostas de ações, programas e estruturas. Para tomadas de decisão, uma avaliação se impõe, qualquer que ela seja. E embora estejamos constantemente avaliando, quando se levanta a necessidade de refletir melhor nas bases dessas decisões, no acompanhamento de sua implementação de uma forma mais sistemática e científica e no uso das informações desse acompanhamento para reforçar ou redirecionar ações num sentido cons-

trutivo, a reação a essa proposta é em geral de desconforto, muitas vezes de desconfiança. Isto depois se traduz sob diferentes formas: numa não integração dos processos avaliativos no cotidiano das decisões, numa manipulação superficial dos dados e até numa negação pura e simples das evidências, quer sejam processuais ou finais, até o ato de simplesmente ignorar o processo de avaliação.

Repensar estas atitudes, recolocar as questões de avaliação sob uma perspectiva mais positiva, parece ser um passo importante a ser dado no processo de democratização da educação, com vistas a uma transparência maior de Políticas e das políticas que dizem respeito ao ensino. Está em jogo aí, também, a nossa competência em fazer acontecer e saber usar.

REFERÊNCIAS BIBLIOGRÁFICAS

- BUCHWEITZ, B.; MOTTA, M. do H. S. (1984). Dupla avaliação de dissertações em concurso vestibular: comparação de escores. *Educação e Seleção*, Fundação Carlos Chagas, nº 10, São Paulo, págs. 83-89.
- Fundação Carlos Chagas. (1987). *Avaliação da educação básica do Nordeste Brasileiro: rendimento escolar*. São Paulo.
- Fundação Cearense de Pesquisa e Cultura (1982). *Avaliação da educação básica no Nordeste Brasileiro*. Relatórios nº 1 a 7, Fortaleza. *Idem*, 1984.
- Fundação Getúlio Vargas, ISOP-CETPP (1971). *Testes de desenvolvimento educacional*, Relatório Técnico, Rio de Janeiro.
- GATTI, B. A. (1977). *Avaliação em sala de aula, Formulação de Objetivos — Avaliação*, Secretaria de Educação do Estado de São Paulo, DRHU-CENP, págs. 40-53.
- GOLDBERG, M. A. A. (1980). A avaliação pode ser uma arma? *Educação e Avaliação*, nº 1, São Paulo, págs. 23-47.
- _____. (1980). Avaliação educacional: medo e poder!!! *Educação e Avaliação*, nº 1, São Paulo, págs. 84-95.
- MELLO, L. de A., LACERDA, W. L. V. (1982). Redação no vestibular: a experiência da Universidade Federal da Paraíba em 1982. *Educação e Seleção*, FCC, nº 6, São Paulo, págs. 27-42.
- RIBEIRO, S. C. e outros. (1981). Flutuação de critérios na avaliação de redações. *Educação e Seleção*, FCC, nº 4, São Paulo, págs. 27-42.
- RIPPER, A. V. (1983). *Análise do vestibular da Fuvest de 1977 a 1981: o efeito das mudanças na sistemática do vestibular sobre a questão das duas fases e os falsos negativos*, Tese de doutoramento, PUC-SP.

- RODRIGUES, C. D.; HAUBOLD, G. T.; LANG, L. M. (1982) Causas de deficiências na expressão escrita em alunos ingressos na Universidade Federal de Santa Maria, *Educação e Seleção*, FCC, n° 5, São Paulo, págs. 45-62.
- SANCHEZ, A. M. T. e outros. (1981). Validade de conteúdo de uma prova de Comunicação e Expressão: análise de alguns problemas. *Educação e Seleção*, FCC, n° 4, São Paulo, 1981, págs. 53-87.
- Secretaria Municipal de Ensino, DEPLAN-DP-SPP (1981). *Verificação do nível de escolaridade dos alunos da rede municipal de ensino*. São Paulo.
- TEIXEIRA, J. G. (1984). Valor real e valor proclamado da prova de redação no concurso vestibular. *Educação e Seleção*, FCC, n° 10, São Paulo, págs. 83-89.
- VIANNA, H. M. (1981). Comunicação e expressão no acesso à Universidade — uma experiência diversificada. *Educação e Seleção*, FCC, n° 4, São Paulo, págs. 45-52.
- _____. (1984). Redação e medida da expressão escrita: algumas contribuições da pesquisa educacional. *Educação e Seleção*, FCC, n° 10, São Paulo, págs. 83-89.
- _____. (1986). *Acesso à Universidade: os caminhos da perplexidade*, Relatório de Pesquisa, FCC, São Paulo, junho.

