

# AVALIAÇÃO DE PROGRAMAS EDUCACIONAIS: DUAS QUESTÕES<sup>1</sup>

*Program and project evaluation is a human activity that is as old as mankind. It is used to make choices and decisions. It motivates change. It is used to develop understanding. It is used to justify investments of resources. In many ways program and project evaluation has been and will continue to be embedded in human cognition and behavior.*

*It is fair to say that every program or project evaluation is unique in many ways and consequently that the practice of evaluation is complex.*

JAMES R. SANDERS<sup>2</sup>

## 1. INTRODUÇÃO

A avaliação educacional, no contexto brasileiro, começou a desenvolver-se tardiamente, em meados dos anos 60, ainda que com a quase total centralização nos processos de medida, situação esta que subsiste nos tempos atuais. Apesar desse relativo desenvolvimento, foi ignorado, por alguns segmentos, o fundamento teórico da avaliação; a subjacente teoria da avaliação em

<sup>1</sup> Artigo publicado na revista *Estudos em Avaliação Educacional*, v. 16, n. 32, p. 43-55, jul./dez. 2005.

<sup>2</sup> In: Kellaghan e Stufflebeam (2003).

que se baseia a ação avaliativa, aspecto este que será ressaltado no presente trabalho. É forçoso reconhecer, entretanto, que já se desenvolveu relativo domínio na avaliação da aprendizagem, na avaliação de competências, na avaliação para fins de seleção e na avaliação de atitudes, começando a ocorrer, no momento fluente, maior domínio da avaliação institucional. Observa-se, quando alguns relatórios são examinados, que a avaliação de programas ainda não está adequadamente definida em todas as suas dimensões, sendo, muitas vezes, confundida com a frequentemente chamada avaliação do desempenho.

A literatura existente sobre avaliação de programas ressaltava diferentes aspectos, como a sua relevância social, as questões técnicas relacionadas com os diversos tipos de validade, a importância da disseminação dos resultados e seus efeitos, entre outras considerações que são abordadas por diferentes autores, inclusive na importante obra editada por Kellaghan e Stufflebeam (2003). Os objetivos do presente artigo limitam-se, entretanto, a duas questões que se referem à natureza do processo de avaliação e ao que se espera dos esforços desenvolvidos em uma avaliação de programa.

## **2. AVALIAÇÃO DE PROGRAMAS: UM PROCESSO DEMOCRÁTICO**

O problema da avaliação como um processo democrático foi discutido com bastante amplitude por House e Howe (2003) e permite considerações relativas a esse assunto nem sempre explorado por teóricos e praticantes da avaliação. O trabalho de avaliar é uma atividade de equipe e, como tal, pressupõe uma ação sem conflitos internos, visando a garantir a validade das conclusões a que a equipe possa chegar na consecução dos seus objetivos. É bom destacar que toda avaliação, inclusive a de programas, tem inserido em seu contexto interesses de diferentes pessoas que refletem diversidade de valores, várias visões de mundo e, conseqüentemente, posicionamentos diversos, traduzidos, por vezes, em situações de conflito, que fragilizam a equipe e perturbam o andamento dos trabalhos. Seria bom lembrar, nesse momento, que, segundo Merleau-Ponty (2004), “... Cada ser é um só, e ninguém pode dispensar os outros... Não há vida em grupo que nos livre do peso

**3** Entende-se como audiências (*stakeholders*, em inglês) todas e quaisquer pessoas envolvidas ou afetadas pela avaliação: estudantes, pais/responsáveis, professores, administradores, orientadores, psicólogos, associações de pais e mestres, futuros empregadores, membros da comunidade e outros que tomem decisões que afetem a educação do estudante.

de nós mesmos, que nos dispense de ter uma opinião...” Fica configurada, dessa forma, a responsabilidade do avaliador na condução de uma equipe de especialistas, com pluralidade de pensares, que procura avaliar um programa.

Assim, como deixam perfeitamente claro House e Howe (2003), é fundamental que se considere quais os interesses, valores e visões das várias audiências<sup>3</sup> em relação ao programa a ser avaliado. Ressalte-se, entretanto, que somente serão levados em conta os interesses, os valores e as visões que são efetivamente importantes para o programa.

A avaliação, conforme o destaque anterior, é um trabalho de grupo, que exige a integração dos seus diversos elementos. Podem ocorrer, entretanto, distorções no pensar e no agir de alguns elementos que impedem a integração do grupo, por ausência de um diálogo fluente e aberto que conduza a decisões sensatas e adequadas à situação. Ainda que difícil, deve-se levar a equipe de avaliação a um consenso, que vai possibilitar ao grupo agir de forma coesa, com a superação de conflitos.

É importante ressaltar o papel do avaliador como elemento aglutinante do grupo, numa avaliação realizada segundo a perspectiva democrática, sendo a ele vedada a sua identificação, ainda que não explícita, com subgrupos que possam vir a surgir na equipe. A ocorrência desses subgrupos deve ser evitada de forma enfática, impondo-se, ainda, por outro lado, que os possíveis desequilíbrios no grupo, gerados por manifestações de autointeresse, sejam igualmente controlados, em respeito ao objetivo maior do projeto.

A avaliação de programas, segundo a perspectiva democrática, pressupõe a participação das várias audiências no desenvolvimento da sua estrutura. Esse aspecto exige, necessariamente, que sejam definidas por antecipação as regras e os procedimentos para esse fim. Isso significa a realização de reuniões prévias, organizadas de acordo com as características das audiências, para que elas possam manifestar suas preocupações e suas ideias relativamente à avaliação e ao respectivo projeto, por via de um processo de interação destituído de formalismos. As audiências, portanto, participam das deliberações relativas ao projeto e, no decorrer dessas reuniões interativas, o avaliador responsável pela equipe procura identificar os reais interesses do grupo, levando-o a uma ação reflexiva. Durante essas reuniões, para fins de deliberação, o avaliador que lidera a equipe pode testar a consis-

tência dos critérios adotados no projeto de avaliação e, ao longo de sua realização, comprovar, igualmente, a coerência dos dados parcialmente coletados, para que, ao término dos trabalhos, sejam válidas as conclusões acerca do programa avaliado.

### **3. ELEMENTOS RESULTANTES DE UMA AVALIAÇÃO DE PROGRAMAS**

A indagação sobre quais seriam os elementos que resultariam das várias ações em uma avaliação de programa admite múltiplas respostas, se considerarmos o grande número de modelos existentes, como pode ser constatado na obra de Madaus *et al.* (1993); desse modo, para responder à questão inicialmente proposta, o presente artigo se baseará em trabalho de Stake (1973), que teve larga repercussão e, decorridos mais de trinta anos, ainda pode servir de orientação para aqueles que se dedicam à avaliação de programas.

A expectativa, segundo o ponto de vista a ser seguido, é de que o relatório da avaliação apresente os objetivos definidos em razão das informações coletadas nas reuniões interativas com as audiências interessadas. Algumas informações são significativas para certos grupos de pessoas, mas essas mesmas informações não atendem aos interesses de outros segmentos das audiências consultadas. O que importa aos responsáveis pela formulação do programa pode não corresponder aos aspectos que são de interesse imediato dos pais de alunos, por exemplo. Assim, é fundamental ter em mente a especificidade dos vários grupos integrantes das audiências, na fase de elaboração de relatórios parciais, no decorrer do processo de avaliação e no momento de estruturação do relatório final.

O documento de disseminação dos resultados da avaliação deve ter um caráter eminentemente descritivo e levar em consideração o público a que se destina. Ele pode não ser de interesse imediato para os que participaram da implementação do programa no seu dia a dia (professores, técnicos escolares, administradores, entre outros), mas ser de grande valia para pesquisadores educacionais e especialistas em avaliação; desse modo, fica definido que na divulgação dos dados de uma avaliação de programa, para que ela tenha impacto, é necessário que a cada tipo de audiência corresponda um relatório específico, variando de um

relatório técnico para especialistas, com as suas complexidades estatísticas, quando for o caso, a um folheto de divulgação dos elementos mais representativos para a sociedade.

Ao ser caracterizada a avaliação como um processo democrático, foi reiterada a necessidade de reuniões interativas com as diversas audiências, ao longo do seu processo. É claro que essas reuniões se devem caracterizar por um intercâmbio de ideias seminais, que fundamentarão o processo de avaliação. A avaliação não é um produto que nasce feito e definido em todas as suas características, ele se constrói com base na troca de pontos de vista, que geram discussões que passam a integrar todo o processo. É igualmente necessário aproveitar a temática desses diversos falares que enriquecem a avaliação e que necessitam ser registrados. Essas discussões, às vezes, são anteriores ao programa e, com grande frequência, outras ocorrem após a sua implantação. O avaliador, nesse caso, ver-se-á obrigado a fazer suposições a respeito do programa, submetendo as suas considerações a rigorosa análise do grupo. As discussões ocorridas constituem um acervo a preservar e são sempre úteis, inclusive no caso de futuras avaliações.

A metodologia de uma avaliação apresenta elementos caracterizados por Stake (1973) como antecedentes que devem ser documentados nos diversos relatórios apresentados às audiências. É necessário que se detalhe a especificação do programa de avaliação, apresentando às audiências os elementos seguintes: 1) o que é o programa e em que consiste; 2) o que ocorreu durante a sua implementação (em sala, em laboratórios ou em outros locais); 3) o que foi tentado, ainda que não tenha sido bem sucedido; e, finalmente, 4) os dados do resultado do programa. Ao relatório incorporam-se também outros importantes subsídios, como: 1) aquilo que se pretendeu e 2) o que foi efetivamente constatado, acrescentando-se outros elementos que, eventualmente, possam ser úteis para a formulação de juízos de valor pela equipe de avaliação. Entre os elementos caracterizados por Stake (1973) como antecedentes, incluem-se ainda: 1) dados demográficos e escolares dos estudantes; 2) características gerais e profissionais dos professores; 3) conteúdos do currículo; 4) material utilizado durante a instrução; 5) descrição física da instituição; 6) organização da escola; e 7) análise do contexto da comunidade em que o programa se desenvolveu.

Ainda dentro da linha apresentada por Stake (1973), a avaliação de programa se preocupa em medir diferentes dimensões, envolvendo habilidades, compreensão e capacidade de interpretar conceitos, por exemplo, cujos resultados devem ser apresentados e interpretados nos relatórios parciais e no documento final. Essas mensurações, ocorridas em diferentes momentos da avaliação do programa, procuram apresentar as suas consequências, ou seja, os seus resultados. Além desses resultados, muitas vezes, no decorrer do projeto, são verificadas atitudes e habilidades motoras. É necessário que se ressalte, quanto às mensurações, que elas não podem ficar restritas ao esquema bastante simplista do pré e pós-teste, devendo, ao contrário, ser adotado um esquema de avaliação formativa.

O relatório de avaliação apresenta, dessa forma, os efeitos do programa em relação aos alunos, porém não fica restrito a esse segmento. Oferece, também, informações relacionadas ao impacto sofrido pelos professores e pela própria instituição em que o programa está sendo aplicado. Essas informações devem ser levantadas por intermédio de múltiplas coletas, com o uso de diferentes tipos de instrumentos, e não podem ficar limitadas à aplicação de um único teste, questionário ou a um simples julgamento de professores, como ocorre em frequentes avaliações.

Ao apresentar o relatório, é necessário que o avaliador responsável pela equipe de trabalho insira uma descrição da filosofia que serviu de base para a definição e estruturação do programa que está sendo objeto da avaliação ou, resumindo, a fundamentação teórica do programa, que não pode ser desconhecida pelos que avaliam, mas é muitas vezes ignorada por quem implementa o projeto. É preciso levar em conta, por outro lado, que existem diferentes posicionamentos filosóficos em relação à educação, sendo necessário que se caracterize, no documento de disseminação da avaliação, qual a filosofia que serviu de base para a definição e a construção do programa.

O avaliador, ao considerar o conjunto das atividades que serão empreendidas pela sua equipe, necessita considerar o corpo do programa na sua inteireza e identificar os aspectos que terão maior destaque pela importância no todo a ser avaliado. Nem tudo é incorporado ao conjunto das preocupações do avaliador. Há aspectos que, por seu significado, configuram realmente o programa e, assim, merecem ser considerados; são

partes que, por serem essenciais ao programa, vão exercer influência no seu êxito.

Stake (1973), com razão, chama a atenção para o fato de que o avaliador não deve procurar identificar, na análise do programa, objetivos referentes a comportamentos. A sua preocupação centra-se, especialmente, nos objetivos da instituição que adotou o programa, nos dos pais/responsáveis em relação ao programa, assim como nos objetivos dos alunos, dos professores e nos objetivos que por ventura outros membros da instituição interessados no programa possam ter. Esses objetivos são considerados durante o transcurso da avaliação, e as várias audiências ficarão a par desses objetivos surgidos aos poucos e identificados no decorrer do processo de avaliação. É preciso ficar claro, portanto, que a avaliação não parte de objetivos, mas os vai identificando no decorrer do processo.

A necessidade de coletar informações ao longo da avaliação é encarecida por Stake (1973), que, entre outras informações, destaca as que se referem 1) a procedimentos instrucionais, 2) a estratégias de ensino adotadas, e 3) a diferentes meios (e multimeios) empregados pelos professores. Esses elementos podem ser levantados por intermédio de entrevistas e pela utilização das múltiplas técnicas de observação (VIANNA, 2000). Ao serem realizadas as observações é necessário que se caracterizem os elementos que comprovam o efetivo desempenho dos professores e dos alunos em sala de aula. Além do mais, é importante que, por intermédio da observação, seja constatada a interação professor/aluno, a ser objeto de análise pela equipe.

Ao realizar a avaliação, além das características ligadas ao nível de competência dos alunos, precisa ser considerada a sua situação social, variável esta que pode concorrer para explicar o desempenho escolar. A situação social, segundo Stake (1973), pode ser apresentada sob a forma de um sociograma<sup>4</sup>, sendo útil esse dado para mostrar os diferentes tipos de contato social, caso se realizem avaliações sucessivas. Outro aspecto a considerar refere-se ao contexto social, ou seja, à relação entre a comunidade e a escola. Ambas devem ser apresentadas com o máximo de detalhes para que, na hipótese de comparações e generalizações, o avaliador possa decidir sobre a efetividade desses procedimentos.

É preciso atentar para o fato de que, ao ser adotado o esquema ora apresentado, com base na orientação de Robert L.

<sup>4</sup> Para uma análise dos métodos sociométricos e sua fundamentação, ver Merleau-Ponty (1990).

Stake, é imprescindível a fixação e definição de padrões para que se possa fazer julgamento de valor sobre a qualidade efetiva do programa. Ao serem definidos esses padrões devem ser levados em conta valores, crenças e múltiplas exigências sociais em relação aos vários tipos de aprendizagem e às diferentes situações da escola. É reconhecidamente sabido por todos que se dedicam à avaliação que esses padrões são difíceis de estabelecer. Assim, tomando-se uma situação hipotética: avaliação de um programa de Matemática ou um outro de Literatura Brasileira para o ensino médio, seria certamente necessária uma consulta significativa para saber a opinião de professores universitários e de professores atuantes nos dois últimos anos do ensino básico, a fim de que seja possível estabelecer padrões de qualidade relevantes para a programação educacional dessas duas áreas curriculares. Ressalte-se, entretanto, que, nessa situação, nem sempre é possível chegar a um resultado consensual, mas o empreendimento precisa ser tentado.

Qual o impacto do programa? Uma avaliação procura determinar em que medida o programa teve algum significado para escola e para o sistema ao possibilitar outras experiências e mudanças de comportamento. Um novo programa, no dizer de Stake, deve oferecer oportunidades para que modificações sociais igualmente ocorram. Por exemplo, qual o impacto da visita dos alunos da 8ª série de uma escola estadual à Bienal do Livro? Qual o impacto das palestras de professores universitários sobre a importância da teoria da relatividade em um novo programa de física para os alunos de uma escola de ensino médio? Qual o impacto provocado nas primeiras séries do ensino fundamental em decorrência da introdução de um programa sobre a arte de contar histórias para professoras? Ou seja, o programa a ser avaliado proporcionou novas oportunidades aos alunos? Houve, realmente, alguma mudança nos alunos, nos professores e na própria escola? Sem dúvida, essas são importantes indagações a serem propostas pelo avaliador ao iniciar o seu trabalho, conforme as colocações de Stake (1973).

Um programa, em princípio, deve proporcionar ganhos aos alunos. Muitas vezes, entretanto, os avaliadores se decepcionam porque os ganhos esperados são reduzidos. Uma primeira explicação para o problema estaria, muito possivelmente, na ausência de instrumentos capazes de constatar diferenças nos diversos

tipos de ganhos no transcurso de um programa. Os instrumentos ora empregados podem ser válidos para a medida de diferentes tipos de atitudes e a identificação de diversos níveis de desempenho, mas não refletem, necessariamente, o impacto dos programas. Essa é uma problemática bastante séria, porquanto, salvo as exceções de sempre, poucas ou raras são as instituições que se dedicam a criar instrumentos sensíveis à medida do impacto provocado pelo desenvolvimento de um programa; consequentemente, como atesta Stake (1973), é difícil apresentar um quadro completo do que está efetivamente ocorrendo no âmbito das instituições educacionais.

A caracterização dos efeitos de um programa é outro aspecto a considerar. Algo acontece, sem sombra de dúvida, nas salas de aula, algo além do tradicional detalhamento dos programas curriculares e das possíveis expectativas de professores, mas esse algo apresenta dificuldades para a tarefa do avaliador, mesmo que seja um profissional bastante experimentado em seu *métier*. Um caminho provável para tentar identificar alguma coisa sobre esse algo desconhecido, que todos sabem existir, segundo a colocação de Stake (1973), estaria em analisar as habituais críticas aos programas educacionais; os comentários, muitas vezes azedos, sobre as metodologias utilizadas; e, ainda, as críticas ao material didático empregado, entre outros aspectos. Uma outra alternativa possível, proposta por Stake (1973), consistiria em fazer uma análise dos dados de diferentes pesquisas educacionais, dados estes que permitirão ao avaliador identificar variáveis úteis para a avaliação e para a solução de seus possíveis problemas.

Todo projeto de avaliação tem um custo que não é apenas financeiro, mas que precisa ser estimado e o seu orçamento controlado, tendo em vista solicitações e novas ideias que possam gerar despesas não previstas. Outros custos existem, como os de pessoal, considerando que uma avaliação de programa exige uma equipe capacitada, cujo custo nem sempre é possível traduzir em termos quantitativos, mas que não deixa de ser um custo, e que a avaliação de programas demanda um tempo razoavelmente longo para a sua concretização. Por outro lado, há o custo relativo ao aluno, que se vê envolvido numa atividade nem sempre interessante para ele. A sua participação significa horas extras de trabalho, aumento de tensão e a consciên-

cia de que está sendo observado por um estranho ao seu grupo, e por mais que se tente explicar que a avaliação é do programa, o aluno sempre acredita ser o objeto imediato da avaliação. Todo esse custo deve ser ponderado, conforme o destaque de Stake (1973), vindo a constituir-se em uma variável cujo valor, no final, nem sempre pode ser estimado com precisão.

Ao ser realizada a avaliação de um programa é imprescindível que se estabeleçam os vínculos que possam existir entre aquilo que foi pretendido e o que foi realmente observado. A análise dessa convergência vai mostrar ao avaliador se pode considerar como aceitável ou não essa ocorrência. Nesse ponto, o avaliador se coloca diante de uma grande interrogação: o que causa realmente o quê? ou, em outras palavras, quais as ações que correspondem a determinados resultados? Fica claro que há necessidade de estudar de que modo as coisas variam simultaneamente, a fim de que se possa estabelecer relações de causa e efeito. A partir dessas relações, comparações podem ser feitas; atente-se, entretanto, para o fato de que a covariância muitas vezes é pouco frequente, ainda que não seja impossível de ocorrer.

Avaliar está associado a julgamentos de valor e sobre isso parece haver razoável consenso, na medida em que é possível chegar a um consenso no campo da educação. Algumas indagações novamente se impõem, como fez Stake (1973): qual o valor do programa? qual o valor das informações coletadas? ou, qual o valor dos objetivos do programa? Ao fazer uma avaliação de programa, o avaliador procura estabelecer, primeiramente, questões de mérito e, a seguir, apresentar os problemas que foram revelados no decorrer do processo de avaliação. Reitere-se, mais uma vez, que na avaliação de programas não se visa a chegar a uma posição consensual, o que é quase impossível, considerando que alunos, professores e administradores podem ter visões inteiramente diferentes sobre um mesmo objeto; assim, a partir de múltiplas fontes, com as mais variadas percepções, o avaliador vê-se na contingência de apresentar juízos de valor que nem sempre coincidem com os que foram manifestados por outras pessoas. A relevância de uma avaliação, conseqüentemente, vai decorrer do bom senso do avaliador, da sua *expertise* e da sua capacidade de interagir com diferentes audiências interessadas<sup>5</sup>.

<sup>5</sup> Para uma visão mais ampla da avaliação de programas em países com larga experiência nessa atividade, ver os artigos de Jean A. King, Alice Dignard e John M. Owen, em Kellaghan e Stufflebeam, 2003, v. II, p.721-768, que tratam da temática, respectivamente, nos Estados Unidos, no Canadá e na Austrália.

#### 4. CONSIDERAÇÕES FINAIS

As partes envolvidas na avaliação de um programa precisam partir, necessariamente, de um processo de negociação, conforme destaque anterior, a fim de que os seus interesses sejam definidos e a avaliação possa atingir os objetivos propostos. A negociação vai possibilitar que os grupos envolvidos, atendendo à dinâmica de um processo democrático, definam suas concordâncias e eliminem possíveis situações de conflito, prejudiciais ao trabalho.

O impacto de um programa de avaliação está diretamente ligado ao sucesso da disseminação dos resultados. Uma política de divulgação dos resultados precisa considerar os diferentes grupos interessados para que, em função de suas características, possam ser selecionadas as informações que correspondem à diversidade dos vários interesses. Assim, a validade de uma avaliação depende grandemente da disseminação criteriosa das informações para as várias audiências.

Ao finalizar o presente artigo, é importante destacar a criação da memória do projeto, atividade que não pode ser desprezada em nenhum momento do seu processo de gerenciamento. As avaliações repetem-se ao longo do tempo e a documentação referente a atividades anteriores pode contribuir para evitar que sejam duplicados trabalhos antes definidos, planejados, desenvolvidos e concretizados. O material para a realização de novas avaliações pode, assim, ser aprimorado à luz de experiências anteriores.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

HOUSE, E. R.; HOWE, K. R. Deliberative democratic evaluation. In: KELLAGHAN, T.; STUFFLEBEAM, D. L. (Ed.) *International handbook of educational evaluation*. Boston: Kluwer Academic, 2003. v. II. p. 79.

KELLAGHAN, T.; STUFFLEBEAM, D. L. (Ed.) *International Handbook of Educational Evaluation*. Boston: Kluwer Academic, 2003. v. II, section 8, p. 699.

MADAUS, G. *et al.* *Evaluation models: viewpoints on education and human services evaluation*. Boston: Kluwer-Nijhoff, 1993.

MERLEAU-PONTY, M. *Conversas: 1948*. São Paulo: Martins Fontes, 2004. p. 50.

\_\_\_\_\_. *Merleau-Ponty na Sorbonne: resumo de cursos, filosofia e linguagem*. Campinas: Papyrus, 1990.

STAKE, Robert L. Evaluation design, instrumentation, data collection, and analyses of data. In: WORTHEN, Blaine R.; SANDERS, James R. (Ed.) *Educational evaluation: theory and practice*. Worthington, Ohio: Charles A. Jones, 1973.

VIANNA, Heraldo M. *Pesquisa em educação: a observação*. Brasília: Plano, 2000.