

Avançando na eficiência de políticas públicas: a avaliação de programas governamentais nos Estados Unidos

Moving forward in the public policies efficiency: the evaluation of governmental programs in the US

Márcio Fernandes PEREIRA¹

Resumo: O objetivo do artigo é realizar uma revisão conceitual da tradição de avaliação de programas governamentais estadunidense. Será dada ênfase aos modos de utilização das diferentes técnicas avaliativas para a verificação dos graus de eficiência dos programas implementados. Inicialmente, será feita uma descrição histórica do desenvolvimento da tradição avaliativa nos Estados Unidos. A segunda seção está devotada aos aspectos conceituais relativos à avaliação de programas. A terceira seção caracteriza o *Modelo Lógico* e de *Teoria do Programa*, estruturas analíticas de auxílio à atividade avaliativa. A quarta seção apresenta distintos atores sociais que podem relacionar-se ao longo da existência de um programa. Finalmente, o trabalho é concluído com a indicação dos limites que envolvem o uso das estratégias avaliativas em programas governamentais.

Palavras chave: Política. Políticas públicas. Avaliação de programas.

Abstract: This article aims to realize a conceptual revision about the tradition of the governmental programs evaluation in the United States. Emphasis will be placed on the ways of use the different evaluative techniques for the verification of degrees of efficiency of the implemented programs. Initially, will be presented a historical description about the development of the evaluative tradition in the United States. The second part is devoted to the conceptual issues related to program evaluations. The third section characterizes the *Logical Model* and the *Program Theory*, analytical frameworks to assist the evaluation activity. The fourth part presents different social actors who can interact during a program existence. Finally, the work is concluded with the indication of the limits that involve the use of evaluative strategies in governmental programs.

Keywords: Policy. Public policies. Program evaluation.

¹ Graduado em Ciência Política pela Universidade de Brasília (UnB) e Mestre em História Econômica Contemporânea (com especialização em Relações Internacionais) no âmbito da Universidade de Coimbra (UC). Pesquisador visitante na Universidade de Salamanca, Espanha. E-mail: <marciofernandes_p@yahoo.com.br>.

Introdução

A avaliação formal de programas governamentais é uma atividade relativamente recente. Os acontecimentos mais significativos para o seu desenvolvimento não são anteriores ao século XX. Por consequência, não é incomum observar entre gestores públicos de países como o Brasil certo desconhecimento quanto a processos sistemáticos de mensuração da eficiência de ações governamentais. Em âmbito nacional, a maior parte das tarefas de acompanhamento da gestão de programas tem se mantido sob a responsabilidade dos tribunais de contas ou de controladorias instaladas em âmbito estadual ou federal, voltadas essencialmente para a fiscalização legal, contábil e orçamentária das ações governamentais. Nestas atividades, a preocupação relativa ao nível de satisfação da população atendida – assim como a definição de diferentes graus de sucesso das ações dos governos – fica secundarizada em função das responsabilidades constitucionais que tais órgãos detêm. Em adição, a tradição avaliativa ainda não encontrou espaço suficientemente adequado em ministérios voltados à gestão de políticas públicas de maior impacto social. Este quadro geralmente aponta para a ineficiência na gestão de programas, o desperdício de recursos, a demora na execução de etapas previstas e mesmo para o desconhecimento do grau mais preciso de sucesso do programa que esta a ser gerenciado.

O objetivo deste artigo é realizar uma breve revisão conceitual da tradição estadunidense de avaliação de programas governamentais. Para tanto, a experiência avaliativa será analisada por meio de um conjunto de trabalhos que, em nossa visão, são capazes de delinear o quadro geral da prática avaliativa nos Estados Unidos, país onde tal prática adquiriu elevados níveis de reconhecimento institucional. Assim, o texto se voltará à exposição dos princípios avaliativos formulados nos estudos de Carol Weiss (1998) e de Worthen, Sanders e Fitzpatrick (2004). Como contribuição à produção daqueles autores, também será levada em consideração a obra das pesquisadoras brasileiras Magdala Raupp e Adriana Reichle (2003).

O artigo está dividido em quatro seções. Inicialmente, será feita uma breve descrição do processo histórico que envolveu o desenvolvimento da tradição avaliativa nos Estados Unidos, dando-se atenção às influências do *Método ZOPP* neste processo. Em seguida, será dada atenção aos aspectos teóricos e conceituais relativos à atividade de avaliação de programas. A terceira seção, por sua vez, está destinada à descrição da ideia de *Modelo Lógico* e de *Teoria do Programa*, estruturas analíticas que auxiliam na identificação das cadeias de causalidade que relacionam as ações dos programas com os objetivos que se deseja alcançar. A quarta seção apresenta uma relação de importantes atores sociais que geralmente interagem entre si ao longo da existência de um programa governamental. Por fim, as

considerações finais salientam os aspectos positivos referentes ao desenvolvimento da tradição avaliativa no setor público, além de indicar limitações desta ferramenta.

1 História da avaliação de programas nos Estados Unidos

A avaliação formal de programas² tem uma história que se relaciona, em seus primeiros momentos, às políticas de educação implementadas em países como o Reino Unido e os Estados Unidos. Neste sentido, os sistemas de inspeções gerais das escolas inglesas do século XIX poderiam ser considerados processos precursores da moderna avaliação. Resultantes de crescentes insatisfações relacionadas aos programas

educacionais daquele período, as ações de inspeção eram realizadas por meio de comissões governamentais que ouviam depoimentos e utilizavam métodos informais para verificar a qualidade de ensino (WATCHTMAN, 1978). Tentativas de se estimar o desempenho de alunos com o objetivo de avaliar a qualidade do ensino também foram realizadas nos Estados Unidos ao longo do século XIX. Tal tendência manteve-se nas primeiras décadas do século XX, com o rápido surgimento de técnicas de mensuração e de testes padronizados baseados nas normas dos programas de educação norte-americanos.

Durante este período, os educadores consideravam as mensurações e a avaliação quase sinônimos, sendo esta última em geral vista como um teste final para avaliar o desempenho de um aluno e atribuir-lhe notas. Embora o conceito mais amplo de avaliação – tal como o conhecemos hoje – fosse embrionário, instrumentos de mensuração úteis para os avaliadores estavam proliferando rapidamente [...]. Publicadas formalmente, avaliações de programas escolares ou currículos viriam a aparecer nos vinte anos seguintes. Uma exceção notável foi o ambicioso *Eight Year Study*, que marcou época ao apresentar um novo modelo de avaliação educacional com sua metodologia sofisticada e sua ligação entre as medidas dos resultados e os resultados desejados da aprendizagem. (WORTHEN; SANDERS; FITZPATRICK, 2004, p. 63).

O projeto avaliativo ligado ao *Eight Year Study*, implementado a partir de 1933, voltou-se para o acompanhamento dos resultados do programa em diferentes escolas de nível médio (WEISS, 1998). Como resultado do *New Deal*, este período também foi marcado pelo

² Nesta etapa, é interessante definir sucintamente o significado da expressão *programas* para que se compreenda o que se quer dizer quando o termo for usado ao longo deste tópico. No âmbito da tradição avaliativa estadunidense, os *programas* podem ser entendidos como um conjunto de elementos organizacionais que, atuando em conjunto, voltar-se-iam para o alcance de metas estabelecidas pelo Governo em relação a uma dada área de sua atuação. Como indicado ao longo de nossa introdução, tais elementos estariam representados, operacionalmente, sob a forma dos recursos, do pessoal gestor e da própria estrutura administrativa necessária às atividades do programa. Por consequência, a *avaliação* se relacionaria à análise das operações e/ou dos resultados de um programa ou política, definindo-se critérios de sucesso ou fracasso como meio de contribuir para o aprimoramento da ação governamental. Nesta atividade, buscar-se-ia manter o rigor investigativo valorizando-se a organização conceitual, a clareza no desenvolvimento de hipóteses explicativas e uma maior atenção quanto à dinâmica social que envolveria o programa. Cf. Weiss (1998).

surgimento de unidades de pesquisa diretamente ligadas a programas implementados pelo governo estadunidense, fomentando a prática da pesquisa social no âmbito dos órgãos públicos daquele país. O conjunto destes programas exigiu estratégias de supervisão voltadas para áreas além da educação. Foram criados órgãos governamentais para acompanhar programas nacionais de assistência social, previdência, saúde, geração de emprego, desenvolvimento urbano.

O período entre as décadas de 1940 e 1960, nos Estados Unidos, foi caracterizado pela expansão da pesquisa social voltada à supervisão de programas. São deste período as primeiras referências a tais trabalhos como *pesquisas de avaliação*, reconhecidas como atividades de investigação voltadas à melhoria de programas governamentais (ROSSI & FREEMAN, 1993). Porém, tais procedimentos ainda se baseavam em metodologias tradicionais no âmbito da pesquisa social. O processo de construção de estratégias avaliativas preocupadas com metas explícitas e mensuráveis iria desenvolver-se no decorrer das décadas seguintes. A definição mais consistente da área que hoje se reconhece como *avaliação de programas* foi desenvolvida ao longo das décadas de 1960 e 1970. Neste período, tiveram papel fundamental para tal processo o planejamento da *Great Society*³ e, mais especificamente, a

³ *Great Society* foi o nome dado a um conjunto de medidas anunciadas em 1964, durante a presidência de Lyndon Johnson (1963-1969), e que objetivava expandir oportunidades de

aprovação da Lei do Ensino Fundamental e Médio (LEFM) de 1965⁴. Diante da necessidade de especialistas em avaliação, o Congresso americano teria aprovado, no final da década de 1960, um importante volume de recursos para avaliações de programas em diferentes setores. Profissionais de áreas ligadas à auditoria, gestão administrativa e planejamento foram procurados. Paralelamente, iniciou-se a publicação regular de trabalhos teóricos e metodológicos ligados à avaliação de programas. O governo federal estimulou o surgimento de programas de pós-graduação na área de avaliação⁵ e

emprego e garantir o crescimento em diferentes setores da economia norte-americana, canalizando recursos para programas em áreas como educação, habitação, geração de emprego e saúde.

⁴ A LEFM previa grandes volumes de financiamento federal para o setor educacional. Tal aporte gerou amplas discussões no Congresso americano, que condicionou sua aprovação à previsão de um igualmente extenso processo avaliativo do projeto. Desta forma, todos os beneficiários do programa teriam que responder a questionários destinados a aferir os resultados dos novos gastos federais. No entanto, a falta de pessoal preparado para executar estas atividades teria se refletido nos péssimos resultados indicados pelo estudo. A concepção tradicional da pesquisa experimental estaria aquém das complexidades do programa. Havia grande dificuldade em identificar as influências que poderiam ser atribuídas aos diferentes componentes da LEFM, bem como separar os efeitos do programa daqueles relacionados a outras atividades realizadas nas escolas. A partir desta experiência, havia ficado clara a necessidade de se construir novas abordagens, estratégias e métodos de avaliação (WORTHEN; SANDERS; FITZPATRICK, 2004).

⁵ Segundo Worthen, Sanders e Fitzpatrick (2004, p. 80), em 1971, mais de 100 universidades norte-

distintos modelos avaliativos foram desenvolvidos. Aos finais da década de 1960, como resultado das experiências adquiridas na gestão de ações dentro do país e do fomento a ações de cooperação exterior, seria desenvolvida a abordagem de *Marco Lógico* (*Logical Framework* ou, simplesmente, logFRAME) no âmbito da *Agência Estadunidense para o Desenvolvimento Internacional* (USAID). Fazendo parte de uma noção geral que já tomava forma na gestão de programas nos Estados Unidos, a abordagem de logFRAME buscava representar as ações programáticas sob a forma de matrizes-síntese em que as diferentes etapas da intervenção eram correlacionadas com atividades intermediárias a serem executadas para a realização da meta final, bem como com os impactos esperados pelo projeto (WORLD BANK, 1996).

A abordagem dos logFRAME acabaria por influenciar outra tradição avaliativa, a alemã, também em formação. Em princípios da década de 1980, baseando-se nas noções de Marco Lógico, a *Sociedade Alemã de Cooperação Técnica* (GTZ) desenvolveria – em parceria com o *Ministério alemão de Cooperação para o Desenvolvimento Econômico* – aspectos próprios de seu método de *Planejamento de Projetos Orientado por Objetivos* (*Zielorientierte Projektplanung*, de sigla

americanas possuíam programas de pós-graduação voltados para a qualificação de avaliadores. Todavia, tal número manteve uma tendência decrescente ao longo do tempo: em 1986, tais cursos restringiam-se a 44 universidades dos Estados Unidos e, em 1993, o número se reduziu para 38.

ZOPP). O valor do método ZOPP estaria relacionado à sua capacidade em adicionar às matrizes de tipo logFRAME diagramas que expusessem os problemas a serem atacados, relacionando-os a objetivos e a possíveis atores sociais envolvidos em sua dinâmica de implementação e funcionamento (BROSE, 1993). Considerando não apenas a etapa de gerenciamento do programa, mas também as fases de desenho e de implementação do projeto, o método ZOPP rapidamente tomaria a forma de uma estratégia participativa e interdisciplinar para o desenvolvimento de programas, levando em conta a visão de distintos públicos e a realização de ações por meio de equipes. Ao longo das décadas seguintes, o ZOPP faria o caminho de volta, influenciando não apenas o processo de desenho e de avaliação de programas nos Estados Unidos, mas também aqueles executados em distintas agências internacionais e países em desenvolvimento. Assim, da relação entre os logFRAME e os métodos orientados por objetivos, derivariam os desenhos de *Teoria do Programa* e os *Modelos Lógicos* mais recentemente utilizados nas atividades avaliativas.

Atualmente, a avaliação de programas ainda é um campo em formação que se relaciona com diferentes disciplinas daquilo que se convencionou chamar de Humanidades. Por envolver processos ligados a estudos de estrutura organizacional e gerencial, por vezes relaciona-se com disciplinas como a administração. Por outro lado, ao se preocupar com aspectos ligados ao melhor uso possível de recursos

escassos, e por envolver a corriqueira intervenção do setor público sobre mercados e processos produtivos, a avaliação não deixa de interagir com a economia. Por muitas vezes abarcar a análise de programas ligados a fenômenos sociais profundos, como a criminalidade, a estabilidade e função dos arranjos familiares e o emprego, tal campo tem se relacionado constantemente com a sociologia e a antropologia. Seu contato com a ciência política não seria menos comum, uma vez que o estudo de programas governamentais vincula-se a importantes núcleos de poder, podendo muitas vezes afetar distintos interesses.

2 Aspectos teóricos da avaliação de programas

2.1 Conceito de avaliação, seus objetivos e definição de programa

O desenvolvimento de práticas avaliativas nos Estados Unidos deu origem a uma série de conceitos com relação à natureza dos programas e de suas partes constituintes. Seguindo tal tradição avaliativa, Weiss (1998) conceitua a avaliação como uma análise sistemática da operação e/ou dos resultados do programa, comparados a um conjunto de critérios explícitos ou implícitos. Seu objetivo seria contribuir para o aprimoramento do programa ou política. Tal análise poderia voltar-se ao entendimento do modo como o programa tem sido conduzido, verificando-se, por exemplo, a correta aplicação de recursos ou o alcance de metas intermediárias. Neste sentido, a

avaliação se voltaria aos aspectos de *processo* que envolvem os programas. De outro modo, a avaliação poderia estar preocupada com os resultados finais e os efeitos do programa sobre seus beneficiários. Desta forma, o foco da análise estaria relacionado aos *resultados* da intervenção governamental.

O julgamento do mérito ou do valor de um programa depende da definição de critérios operacionais previamente estabelecidos. Estes critérios podem ser originários das próprias expectativas dos gestores do programa, ao definirem oficialmente metas a serem atingidas. Tais critérios seriam, por consequência, explícitos. De outro modo, não é incomum a ausência de definições claras quanto aos fins esperados pela ação governamental. Neste caso, caberia ao avaliador buscar o maior conhecimento possível sobre as atividades que envolvem o programa analisado. Este tipo de informação pode ser importante para a identificação de possíveis critérios implícitos contidos na dinâmica organizacional. Mesmo as metas oficiais podem ser alteradas ao longo do tempo em virtude dos aspectos práticos que envolvem a intervenção dos agentes públicos. De outra forma, podem ser identificados elementos de julgamento provenientes de setores não pertencentes à equipe do programa. Como salientam as abordagens ZOPP, as expectativas de financiadores, tomadores de decisão ou mesmo da população beneficiária podem acabar tornando-se critérios de avaliação.

Não há unanimidade quanto ao que os avaliadores pensam ser o objetivo

intrínseco da avaliação. Para Weiss (1998), a atividade tem o objetivo de contribuir para o aprimoramento do programa ou, mais especificamente, colaborar para a melhoria de sua qualidade no que diz respeito à capacidade de atingir as metas para as quais ele foi criado. Worthen; Sanders; Fitzpatrick (2004), por sua vez, buscam estabelecer distinção entre aquilo que eles chamam de *objetivo* e *papéis* da avaliação. Para eles, o objetivo básico da avaliação seria produzir julgamentos de valor quanto aos programas a serem estudados. Tendo estabelecido suas considerações a respeito do programa, a avaliação poderia ser utilizada para diversos fins (por exemplo, ajudar gestores a tomarem decisões e definirem metas a serem atingidas, ou garantir a melhor alocação possível de recursos de uma ação para outra). Esses seriam seus *papéis*. Concordando com Scriven e Donaldson (2002), os autores salientam que "[...] a meta da avaliação é dar respostas a perguntas avaliatórias significativas que são apresentadas, ao passo que os papéis da avaliação se referem às maneiras pelas quais essas respostas são usadas" (WORTHEN SANDERS; FITZPATRICK, 2004, p. 39).

A atividade de avaliação que Weiss (1998) se propõe a apresentar tem como objeto principal os próprios programas. Porém, como a autora salienta, estes não são os únicos elementos passíveis de estudo. Weiss faz distinção entre *política*, *programa*, *projetos* e *componentes*, todos passíveis de avaliação. *Política* (*policy*) seria a posição oficial de um governo quanto a um conjunto de objetivos a

serem perseguidos em uma área específica. Desta forma, faria sentido se falar em *política ambiental*, *política de emprego* ou *política agrícola* de um governo, tendo cada uma destas áreas um conjunto específico de programas. Apesar de não haver uma definição consensual quanto ao significado do termo *programa*, sua designação geralmente diz respeito a um conjunto de pessoas, organizações e recursos mobilizados de maneira deliberada e contínua para alcançar um objetivo pré-determinado. Também é comum dividir-se as ações de um programa em dois elementos principais – *atividades* e *projetos*. Os *projetos* seriam tarefas específicas dentro de um programa, com caráter de investimento, planejadas para serem executadas com início, meio e fim, não importando a sua abrangência. As *atividades* seriam tarefas rotineiras dentro do programa, tendo apenas caráter de manutenção do mesmo, sendo executadas de maneira contínua e regular (ROYSE, 1992).

A tradição avaliativa alemã reforçou a noção de que os programas podem ser avaliados em suas distintas fases de existência. Raupp e Reichle (2003) dividem-nos em quatro fases: *diagnóstico das necessidades*, *desenho* (*formulação*), *implantação* e *gerenciamento de atividades*. A fase de diagnóstico estaria ligada à identificação dos problemas a serem resolvidos no âmbito da dinâmica social. A etapa de desenho seria aquela relacionada à formulação do programa no que diz respeito à sua estrutura organizacional, atividades e ações, bem como à sua margem de atuação e

recursos. A fase de implantação se relacionaria com a colocação em operação dos elementos organizados no decorrer da etapa do desenho. Por fim, o gerenciamento se vincularia ao processo de administração do programa ao longo de sua existência.

Porém, um programa não é necessariamente desenvolvido para ter início, meio e fim. Não é incomum a existência de programas de caráter continuado, desenvolvidos sem a previsão formal de seu término. Isso não significa que eles não possam passar por sucessivas modificações. O fato é que poucos programas são tão bem feitos a ponto de não necessitarem de alterações no decorrer de sua operação. Por outro lado, a extinção de um programa governamental é um processo difícil do ponto de vista político. Uma vez iniciados, os programas passam a ter vida própria. Suas ações desenvolvem clientelas que usufruem dos serviços do programa e que dificilmente teriam o interesse em vê-lo desaparecer. A eliminação de certos programas, principalmente os de grande abrangência, geralmente tem um alto custo político para os tomadores de decisão. Desta forma, alterações na órbita dos programas podem ser mais comuns que extinções, indicando os riscos de se ignorar a dinâmica social em torno de um programa ao se tentar compreender a manutenção de ações aparentemente ineficientes.

2.2. Tipos de avaliação

Inicialmente formulada por Scriven (1967), a distinção entre avaliações *formativa* e *somativa* está relacionada às intenções do avaliador ao realizar o estudo. Neste aspecto, a *avaliação formativa* seria feita para disponibilizar informações à equipe gestora do programa, responsável por aprimorá-lo ou torná-lo mais eficiente. Este tipo de avaliação estaria ligado a decisões quanto ao desenvolvimento do programa, como sua modificação ou revisão. De outro modo, a *avaliação somativa* seria realizada para oferecer aos tomadores de decisão julgamentos de valor ou mérito quanto à existência do programa. Tal avaliação estaria vinculada a decisões sobre a continuação do programa, seu encerramento, sua ampliação ou adoção.

Deve-se tomar cuidado para não se confundir avaliações *formativas* ou *somativas* com avaliações de *processo* ou de *resultado*. De fato, as avaliações de processo examinam o que tem acontecido no decorrer da implantação e da operação do programa, enfatizando aspectos como envolvimento dos participantes, tarefas executadas, desempenho dos agentes estatais, etc. Avaliações de resultado, por sua vez, focalizam os impactos da ação do programa sobre os beneficiários. Todavia, não se deve perder de vista que o caráter formativo ou somativo relaciona-se às *intenções* do avaliador e à *natureza da decisão* a ser tomada. De outra forma, "[...] *processo* e *resultado* não têm nada que ver com o papel do avaliador, mas antes se relacionam com a *fase* em

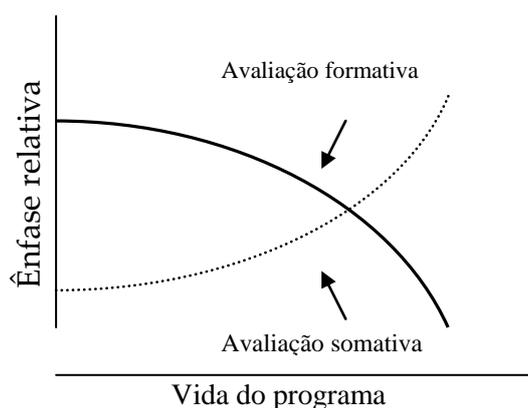
que o programa está sendo estudado" (WEISS, 1998, p. 32, tradução nossa).

As avaliações formativas tendem a ser feitas nas fases iniciais da operação de um programa, enquanto que as avaliações somativas são executadas em seus estágios finais. Mas tal observação não é restritiva. Decisões quanto à continuidade ou encerramento do programa podem ser tomadas em suas

fases iniciais, diante de problemas inesperados, e exigências de aprimoramento quanto a algum aspecto da intervenção governamental podem ser necessárias nas suas últimas etapas. Porém, como reconhecem Worthen Sanders; Fitzpatrick (2004), a tendência para se realizar avaliações formativas ou somativas muda ao longo da vida de um programa:

Figura 1

Relação entre a avaliação formativa e a somativa durante a vida de um programa



Fonte: Worthenet al.(2004, p. 49).

Outras distinções podem ser feitas quanto à tipologia dos estudos de avaliação. Raupp e Reichle (2003) fazem menção a mais duas atividades avaliativas: a *avaliação diagnóstica* e o *monitoramento*. A avaliação diagnóstica seria realizada antes do desenho de um programa. Seu objetivo seria identificar um possível problema, indicando sua natureza e prováveis causas, ajudando na formulação de um programa futuro. Tal avaliação representa a tentativa de se formular um programa de acordo com a dinâmica social e política sobre a qual ele

atuaria. O monitoramento, por sua vez, seria uma atividade rotineira no âmbito de funcionamento do programa. Sua execução estaria voltada a verificar se tarefas planejadas foram realmente executadas. O monitoramento faria uso de registros administrativos para acompanhar a alocação de recursos humanos, materiais e financeiros, bem como para observar a correta obediência ao cronograma previsto.

2.3. Desenhos avaliativos

O desenho da avaliação designa o conjunto de procedimentos previamente definidos que serão utilizados para se verificar o mérito do programa. De certa forma, o desenho indica "[...] quais pessoas ou unidades serão estudadas e como elas serão selecionadas, quais tipos de comparações serão feitas e o tempo de duração da investigação" (WEISS, 1998, p. 180, *tradução nossa*). Exigências como definição de variáveis, indicação da população a ser estudada, período de realização da pesquisa, intervalo temporal considerado, capacidade de extensão das conclusões fazem parte do planejamento do desenho avaliativo.

A lógica da avaliação está ligada a dois elementos: a) a comparação da situação dos participantes do programa *antes* e *depois* de receberem os benefícios da intervenção, com o objetivo de verificar se houve diferença de ganhos com base em indicadores pré-definidos; b) a comparação dos *participantes do programa* com um grupo equivalente de *pessoas que não participaram* do programa, com o objetivo de verificar se os ganhos dos beneficiários excederam aqueles recebidos pelos não-beneficiários, em um dado período, com base em indicadores pré-definidos (WEISS, 1998). Quanto ao segundo aspecto, preferencialmente opta-se pela seleção de um *grupo de controle*.

O grupo de controle seria selecionado dentro da mesma população de origem dos participantes do programa, com a exceção da amostra ser composta por indivíduos que não foram beneficiários da ação governamental. Para tanto,

pode-se fazer uso de técnicas de amostragem aleatória – a avaliação teria assim um caráter *experimental*. Os grupos sofreriam as mesmas influências externas, desenvolver-se-iam durante o mesmo intervalo de tempo e seriam submetidos aos mesmos testes. A diferença entre ambos se restringiria ao fato de um grupo ter sido exposto ao programa enquanto o outro, não.

Porém, quando a seleção aleatória não é possível, opta-se por um processo de escolha dos não-participantes com base em critérios mais ou menos arbitrários, buscando-se certa comparabilidade entre beneficiários e não-beneficiários. Neste caso, a avaliação seria denominada como *quase-experimental*. Nestas avaliações, a busca por um *grupo de comparação* – termo utilizado para designar grupos selecionados por procedimentos não-aleatórios – é feita após a definição dos critérios básicos que compõem o conjunto de participantes do programa. Os desenhos quase-experimentais possuem limitações claras quanto à possibilidade de generalização de resultados. Sua função é servir de abordagem inicial de um aspecto do programa, dando contribuição limitada ao estudo de seu funcionamento e eficiência.

A realização de processos avaliativos com base em dois grupos comparáveis nem sempre é possível. Quando isso acontece, pode-se optar por procedimentos que Weiss chama de *desenhos de grupo único* (*one-group designs*). Neste tipo de desenho, apenas os participantes são considerados. Tais

desenhos são divididos em duas categorias: *somente depois* (*after only*) e *antes e depois* (*before and after*). Abordagens *somente depois* são realizadas quando a avaliação é executada depois do programa já ter sido implantado. Desta forma, a análise é feita por meio do auxílio de registros administrativos do programa, relatórios de monitoramento das tarefas realizadas, entrevistas com atores importantes para a execução do programa. Pode-se também realizar a comparação da situação dos participantes inseridos posteriormente no programa com aqueles beneficiários mais antigos. Neste tipo de avaliação, a existência de documentação ligada ao programa é especialmente importante.

As avaliações *antes e depois* são caracterizadas pela presença do avaliador antes mesmo da implantação do programa, quando sequer existem registros administrativos de suas tarefas. A ideia seria acompanhar os beneficiários desde o momento em que eles entram no programa até a sua saída. Tal procedimento exige a consideração de limitações. Enquanto o programa está em operação, uma série de acontecimentos não diretamente vinculados e ele podem afetar a vida dos beneficiários. Uma dada situação social sempre é produto de uma interação muito grande de variáveis, cada uma delas mais ou menos importantes para o resultado final. Assim, cabe ao avaliador levar em conta o contexto sobre o qual o programa atuou. A utilidade destas avaliações seria antes de tudo *formativa*, garantindo uma verificação preliminar quanto aos impactos de um programa.

Para fins *somativos*, procedimentos mais rigorosos de avaliação experimental devem ser considerados.

Desenhos de *grupo único* também poderiam ser elaborados com maior extensão temporal, buscando coletar dados muito antes e muito depois da finalização dos programas, ou mesmo em diferentes etapas de seu gerenciamento. Tal orientação visa determinar a possibilidade de auto-sustentação do programa, sua capacidade de manutenção dos resultados benéficos de suas ações ao longo do tempo, mesmo após o fim de suas atividades. Estudos seriais também poderiam indicar se "medidas imediatamente antes e após o programa são uma continuação de padrões anteriores ou se elas marcam mudanças decisivas" (WEISS, 1998, p. 197, *tradução nossa*). Outros procedimentos poderiam ser utilizados para substituir a incapacidade de comparação dos desenhos de *grupo único*. Pode-se, por exemplo, adotar abordagens de *reação a doses* (*dose-response designs*). Nelas, o avaliador compararia participantes que receberam grandes quantidades de bens ou serviços prestados pelo programa com aqueles que receberam menores quantidades. O objetivo seria identificar se maiores quantidades do benefício prestado gerariam melhores resultados quanto aos critérios definidos pelo programa.

3 Teoria do programa e modelo lógico

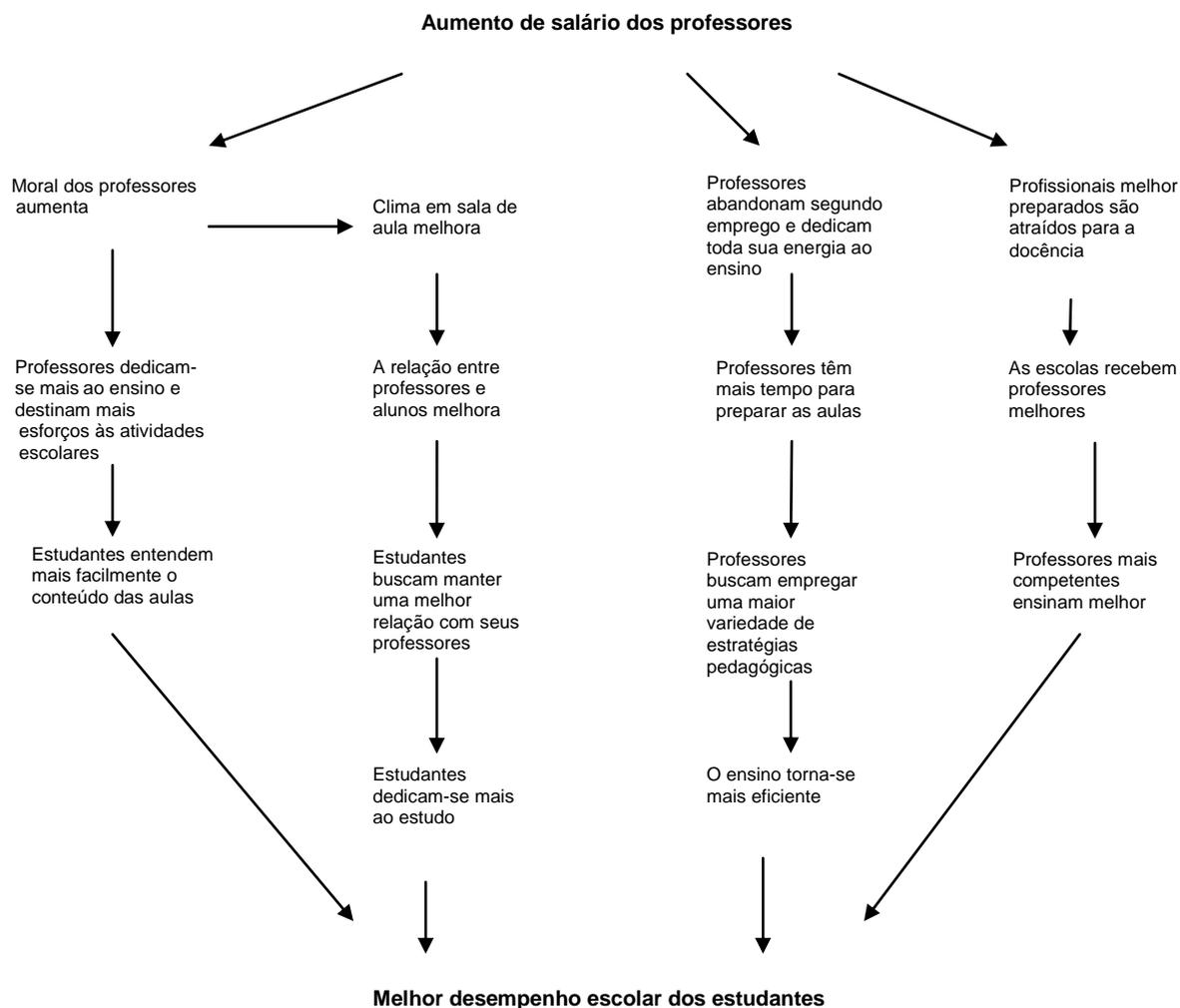
A atividade avaliativa requer um conhecimento consistente quanto aos

mecanismos e à lógica de funcionamento do programa. Como visto, tal preocupação sempre esteve presente nos processos de desenvolvimento de técnicas avaliativas, tomando formas amplamente reconhecidas por meio das abordagens logFRAME e ZOPP. Como um conjunto de recursos, pessoas e organizações, os programas podem envolver relações complexas, de difícil compreensão quanto a execução e funcionamento. Para facilitar o entendimento da lógica que orienta as ações de um determinado programa, Weiss (1998) sugere o desenvolvimento das *Teorias do Programa*.

Herdeira das abordagens logFRAME, a Teoria do Programa representa a conexão lógica referente aos impactos causados pela intervenção sob análise. Ela refletiria a percepção dos

formuladores, gestores e mesmo de outros interessados quanto ao conjunto e à natureza das ações que, combinadas, gerariam um resultado desejado. A expectativa quanto a tal reação é geralmente representada sob a forma de um diagrama de causalidade em que diferentes variáveis influenciam umas às outras, gerando determinado resultado ao final da interação. Para facilitar seu entendimento, Weiss (1998) exemplifica a construção do diagrama com base em um programa hipotético de aumento salarial voltado a professores. De acordo com o exemplo, a justificativa para o aumento se relacionaria ao impacto de tal medida na melhoria do desempenho escolar dos alunos. A autora apresenta diferentes possibilidades quanto à reação dos beneficiários diante do impacto do programa:

Figura 2
Alguns mecanismos por meio dos quais a melhoria do desempenho escolar dos alunos pode ser ligada ao aumento de salário dos professores



Fonte: Weiss(1998, p. 56, tradução nossa).

A autora salienta que a Teoria do Programa não precisa necessariamente indicar impressões corretas quanto às relações de causalidade. Como mencionado, ela reflete as percepções de gestores e formuladores com relação às respostas sociais diante dos estímulos gerados pelo programa. A Teoria do

Programa tem propósitos orientadores em relação à verificação de *falhas da teoria* e *falhas de implantação*⁶. Assim, a Teoria

⁶ As *falhas da teoria* indicam as razões pelas quais um programa não obteve o sucesso esperado, pela não efetivação das expectativas de causa e efeito contidas na sua Teoria do Programa. As *falhas de implantação*, por sua vez, também indicam as razões de fracasso do programa,

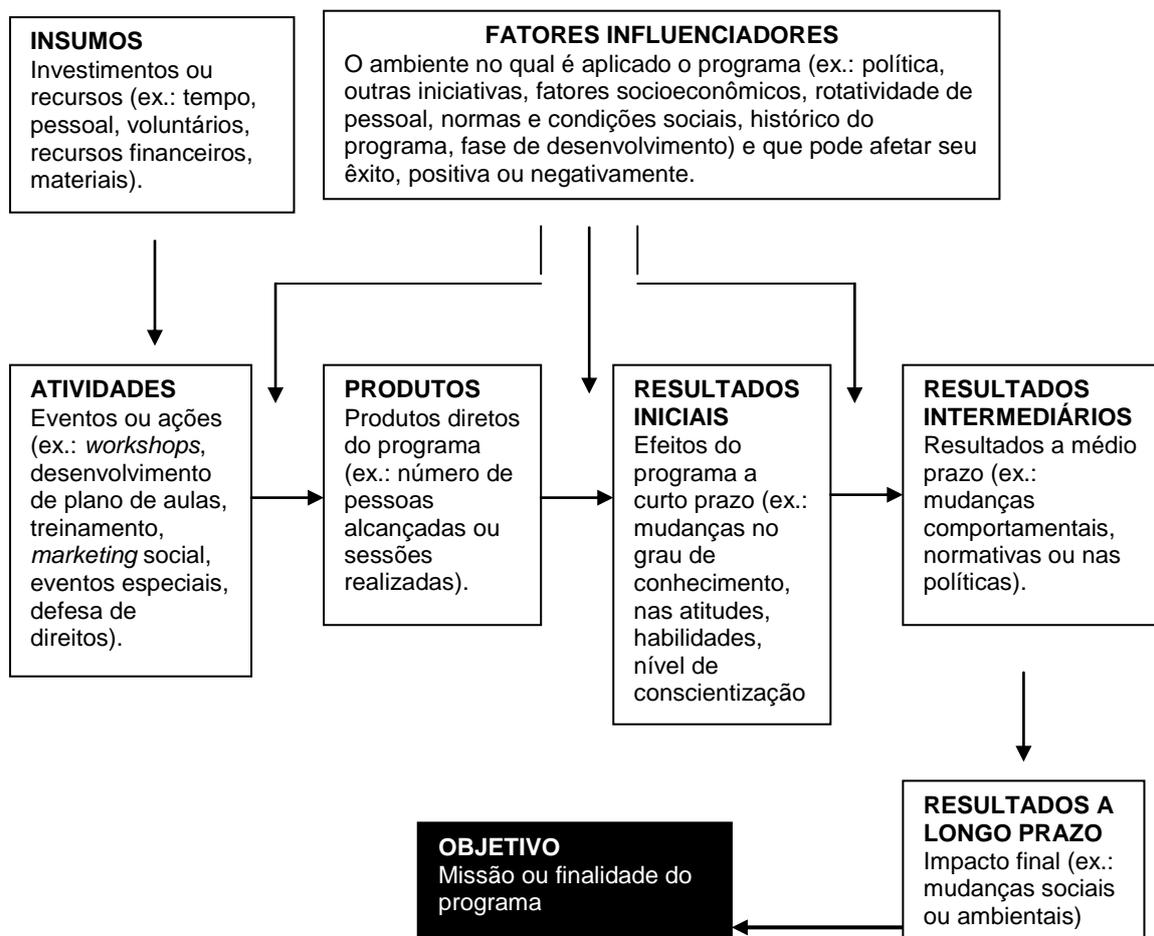
do Programa pode indicar a natureza da complexidade da intervenção governamental, sugerir pontos de risco quanto ao alcance de resultados e orientar ações avaliativas.

Modelo Lógico qualquer pode ser representado da seguinte maneira:

A Teoria do Programa pode ser auxiliada em sua função pelo *Modelo Lógico*. O Modelo Lógico é um diagrama que caracteriza o conjunto de atividades e resultados que pautam o programa no que diz respeito aos seus diferentes recursos. Trata-se de um modelo esquemático claramente influenciado pela proposta ZOPP e que descreve os componentes essenciais do programa, indicando as suas conexões com os resultados esperados. O Modelo Lógico é comumente caracterizado por quatro conjuntos de elementos: *insumos*, *atividades*, *produtos* e *resultados*. Os insumos dizem respeito aos recursos aplicados durante a realização do programa. As atividades indicam aquilo que o programa realizaria ao fazer uso dos insumos, com o intuito de cumprir com seus objetivos. Os produtos representam os elementos finais, acabados, derivados da execução das diversas atividades programáticas. Os resultados, por sua vez, representam os benefícios ou mudanças gerados pelo programa, normalmente em função da conjunção desta cadeia de elementos. O Modelo Lógico também pode levar em consideração o contexto social no qual se insere o programa. O diagrama de um

apontando antes para os problemas ligados à não realização de atividades ou à não execução de recursos que, por algum motivo, não puderam ser acionados no tempo correto. Cf. Weiss (1998).

Figura 3
Componentes de um Modelo Lógico Básico



Fonte: UnitedStates(2002, p. 17).

A relação entre a Teoria do Programa e o Modelo Lógico possibilita a identificação das atividades programáticas que podem estar relacionadas a produtos ou a resultados específicos, valorizando a relação de causalidade entre eles. Desta maneira, os avanços progressivos do programa podem ser comparados às previsões em relação aos diferentes impactos esperados pela intervenção governamental. Porém, é também este caráter esquemático, segundo alguns autores (GASPER, 1999), um elemento de

fraqueza destas abordagens. Análises de longo prazo ou que envolvam a implantação simultânea de ações de mesmo desenho em regiões de geografia e população fortemente distintas geralmente acusam a fragilidade analítica de Teorias do Programa e de Modelos Lógicos. Usados de forma isolada, tais instrumentos podem sujeitar-se a uma notável rigidez incapaz de captar todas as informações derivadas da relação entre populações beneficiárias, ambiente e instrumentos

programáticos. Em adição, a forte dependência às noções de matriz lógica – tão comum às abordagens logFRAME e ZOPP – podem enviesar conclusões e desconsiderar sutilezas que só podem ser percebidas mediante abordagens qualitativas de forte preocupação simbólica. Não se pode esquecer de que os ambientes sobre os quais atuam os programas governamentais também são espaços de interação social, de relações de poder e de dinâmica valorativa profunda nos quais também estão inseridos os avaliadores e seus instrumentos de análise.

4 Atores envolvidos com o programa e dinâmica social

Pela própria natureza de sua dinâmica, os programas governamentais constituem sistemas políticos⁷ caracterizados pela interação de diferentes atores sociais. A correlação entre recursos e estruturas organizacionais, somadas a interesses particulares, faz dos programas um espaço de interação entre distintas forças políticas (BEHRING & BOSCHETTI, 2006). Neste sentido, a literatura ligada à

⁷ Por *sistema político*, entende-se "[...] qualquer conjunto de instituições, grupos ou processos políticos que têm certo grau de interdependência recíproca" (URBANI, 2000, p. 1163). O ponto originário de sistema político seria a noção de que o homem está preso a uma multiplicidade de interações, por meio das quais condiciona os demais e é por eles condicionado, formando um complexo de relações. Assim, qualquer agrupamento poderia ser analisado sob um duplo aspecto: como um conjunto de indivíduos e como uma "rede de relações". Para mais detalhes, Cf. Urbani (2000).

avaliação normalmente designa sob o termo de *stakeholders* os grupos de indivíduos que, de alguma forma, são interessados ou afetados pelas ações de um programa ou pelos resultados de sua avaliação (WEISS, 1984; WORTHEN SANDERS; FITZPATRICK, 2004). O termo também inclui aquelas pessoas que formulam, implementam, gerenciam ou tomam decisões quanto à intervenção governamental em questão.

De modo operacional, os *stakeholders* podem ser divididos entre *implementadores*, *parceiros*, *participantes* e *tomadores de decisão* (UNITED STATES, 2002, p. 11). Os *implementadores* são aquelas pessoas envolvidas nas operações do programa. Podem ser divididos entre *formuladores* – quando fazem parte da equipe de criação do programa – ou *gestores* – quando integram a equipe de gestão do programa. Tal distinção é importante, pois nem sempre o grupo de pessoas responsável pela criação do programa irá constituir a equipe designada para operá-lo. Os *parceiros* integram aqueles atores que apóiam ativamente ou se interessam pelas ações realizadas pelo programa. Podem ser identificados sob a forma de órgãos auxiliares à sua execução ou grupos organizados, de caráter público ou privado, interessados em sua continuidade. Em contraponto aos *parceiros*, também não é incomum a identificação de atores que se opõem à execução do programa.

Os *participantes* – ou *beneficiários* –, por sua vez, integram aquelas pessoas atendidas ou afetadas pelas ações

avaliadas. Sua influência sobre o funcionamento dos programas governamentais pode ser notável e as percepções de tais grupos quanto à natureza da intervenção pode indicar aspectos avaliativos importantes. Por fim, os *tomadores de decisão*, como a designação indica, compõem aqueles que têm poder para tomar decisões quanto à continuidade, alteração ou ampliação do programa. Geralmente, são compostos por autoridades políticas ou administrativas de destaque. É interessante notar que tal classificação estabelece uma distinção entre *tomadores de decisão* e *implementadores*. De fato, nem sempre tais funções estão integradas na órbita de um mesmo grupo, fazendo com que, muitas vezes, aqueles responsáveis pelo gerenciamento do programa tenham pouca liberdade quanto a alterações formais em sua lógica de funcionamento. Mais uma vez, a diversidade de atores sociais envolvidos nas atividades de execução de um programa atenta para a necessária atenção à dinâmica social relativa às ações governamentais.

Considerações finais

A ausência de estratégias avaliativas nas etapas de formulação e de gerenciamento de programas acaba por contribuir para a implementação de ações governamentais carentes de uma clara consciência quanto ao grau de sucesso das medidas adotadas. O resultado é o estímulo à formação de programas *incrementais*, que sofrem sucessivas alterações improvisadas ao longo de sua existência, tornando-os

estruturas cada vez mais complexas e custosas. Ao longo deste processo, o programa exige maiores gastos, sendo incapacitado de atingir o seu máximo potencial. Procedimentos incrementais tendem a prolongar medidas que deveriam ter curta duração, a atrasar obras que antes possuíam prazo certo e a aumentar previsões orçamentárias para a execução de empreendimentos. Tais problemas não constituem novidade no Brasil e a falta de eficiência na gestão de políticas públicas é elemento importante para a compreensão desta dificuldade em avançar na capacidade de cumprimento de metas. A despeito das críticas que se possa fazer às técnicas de avaliação internacionalmente utilizadas, o fortalecimento da tradição avaliativa no país certamente contribuiria para um aumento na capacidade governamental de corrigir erros e alcançar objetivos. De qualquer modo, deve-se salientar o contexto social em relação ao qual as avaliações tendem a ser feitas e, assim, não se esperar mais do que aquilo que elas podem oferecer ao gestor de políticas públicas.

Um elemento que não pode ser desconsiderado nos estudos de avaliação diz respeito ao fato das avaliações de programas desenvolverem-se em um contexto político. Decisões decorrentes do exercício político podem não coincidir com uma análise mais formal de custo-benefício. Políticos profissionais podem, por exemplo, estar interessados principalmente em programas que beneficiem grupos eleitorais específicos, capazes de apoiá-los ao longo dos diferentes ciclos eletivos. Desta forma, o

interesse pelo programa poderia manter-se independentemente dele causar ou não impacto real sobre problemas sociais a serem mitigados.

Como mencionado, raramente os programas são mantidos exatamente da mesma forma como foram criados. Eles estão mudando constantemente. Esta mudança pode ser um efeito direto de processos avaliativos, mas frequentemente é resultado de alterações no ambiente social. Quando um programa entra em operação, mesmo os seus objetivos reais podem mudar. Tais modificações podem ser positivas, garantindo a adaptação do programa às condições reais do ambiente em que ele está sendo gerenciado. Porém, diversos interesses políticos podem antes estar preocupados com a manutenção do orçamento de certos programas do que com as melhorias das ações da intervenção governamental. De fato, as organizações governamentais e seus programas podem servir a múltiplos grupos e interesses, cada um deles com diferentes expectativas com relação aos impactos das intervenções públicas.

Mesmo a atividade de avaliação pode ser influenciada por limitações empíricas. Por exemplo, uma vez que os objetivos de um programa foram identificados e expressos de forma clara, o próximo passo seria desenvolver um meio para medir a extensão em que tais metas foram alcançadas. Porém, a mensuração de resultados pode ser difícil no setor público. Um problema fundamental é a ausência de técnicas pré-existentes de julgamento do valor daquilo que foi

efetivamente realizado. Outros fatores inibem uma adequada mensuração do desempenho governamental. Um deles é o lapso de tempo para que os impactos do programa sejam perceptíveis. Alguns programas geram efeitos somente após anos de implantação, outros têm resultados quase imediatos, mas por pouco tempo. A mensuração dos efeitos de um programa também pode ser afetada pela confusão entre o seu processo de implantação e a própria história das tentativas anteriores de resolução do problema. O ato de avaliação é naturalmente limitado e pode não conseguir captar todas as faces dos programas governamentais. Por tais motivos, conhecer e compreender o processo de desenvolvimento dos problemas para os quais o programa foi formulado é uma etapa importante da avaliação, contribuindo para um uso mais maduro desta ferramenta capaz de aumentar a eficiência de políticas públicas.

Referências

BEHRING, E.; BOSCHETTI, I. **Política social: fundamentos e história**. São Paulo: Cortez, 2006.

BROSE, M. **Introdução à Moderação e ao Método ZOPP**. Recife: Deutsche Gesellschaft für Technische Zusammenarbeit: GTZ, 1993.

GASPER, D. Problems in the Logical Framework Approach and Challenges for Project Cycle Planning. **The Courier Review**, n. 173, 1999.

- RAUPP, M.; REICHLE, A. **Avaliação:** ferramentas para melhores projetos. Santa Cruz do Sul: EDUNISC, 2003.
- ROSSI, P. H.; FREEMAN, H. E. **Evaluation:** a systematic approach. Newbury Park: Sage, 1993.
- ROYSE, D. **Program Evaluation.** Chicago: Nelson-Hall, 1992.
- SCRIVEN, M. The Methodology of evaluation. In: STAKE, R. E. (Ed.). **Curriculum Evaluation - American Educational Research Association Monograph Series on Evaluation**, No. 1. pp. 39-83. Chicago: Rand McNally, 1967.
- SCRIVEN, M.; DONALDSON, S. **Evaluating Social Programs and Problems:** visions for the new millennium. Mahwah: Lawrence Erlbaum Publishers, 2002.
- SCRIVEN, M. Meta-Evaluation Revisited. **Journal of MultiDisciplinary Evaluation**, v. 6, n. 11, Jan. 2009.
- UNITED STATES. Department of Health and Human Services. **Physical activity evaluation handbook.** Atlanta: Centers for Disease Control and Prevention, 2002.
- URBANI, G. Sistema político. In: BOBBIO, N.; MATTEUCI, N.; PASQUINO, G. (Orgs.). **Dicionário de política.** Brasília: EdUnb, 2000. v.2.
- WATCHMAN, E. L. Evaluation as a story: the narrative quality of educational evaluation. In: AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, March, 1978. Toronto. **Annals of the American Educational Research Association, Toronto, 1978, March.**
- WEISS, C. H. **Using Social Research in Public Policy Making.** Lexington: Lexington Books, 1977.
- WEISS, C. H. Toward the future of stakeholder approaches in evaluation. In: CONNER, D.; ALTMAN, G.; JACKSON, C. (Eds.). **Evaluation Studies Review Annual**, Vol. 9, Beverly Hills: Sage, 1984.
- WEISS, C. H. **Evaluation:** methods for studying programs and policies. New Jersey: Prentice Hall, 1998.
- WORLD BANK. **The World Bank Participation Source-Book.** Washington D.C.: Environmentally Sustainable Development-World Bank, 1996.
- WORTHEN, B.; SANDERS, J.; FITZPATRICK, J.. **Avaliação de Programas:** concepções e práticas. São Paulo: Gente, 2004.