MINISTÉRIO DA EDUCAÇÃO
Secretaria de Educação Continuada, Alfabetização e Diversidade
Departamento de Avaliação e Informações Educacionais

Texto Referencial para Avaliação, Monitoramento e Acompanhamento– Caderno de Estudos Temáticos

Introdução

Este texto representa uma iniciativa do Departamento de Avaliação e Informações Educacionais da Secretaria de Educação Continuada, Alfabetização e Diversidade de explicitação do marco conceitual que orienta o trabalho desta unidade. Pretende-se, por este meio, contribuir para a melhor compreensão das funções e forma de atuação deste Departamento e, ainda, para a disseminação das práticas de avaliação, acompanhamento e monitoramento de programas no âmbito da SECAD.

Na primeira seção, situa-se a avaliação como etapa do ciclo de gestão que tem ganhado primazia a partir da década de 1990, num contexto de reforma do Estado e de busca por uma gestão orientada por resultados. Esta seção discute ainda como a avaliação, o acompanhamento e o monitoramento têm se inserido no âmbito da administração pública federal e apresenta algumas das iniciativas existentes, destacando a localização institucional das áreas de avaliação e monitoramento e a interlocução dos Ministérios setoriais com os órgãos de planejamento e de controle interno e externo sobre avaliação.

A segunda seção dedica-se a definir conceitos básicos adotados pelo DAIE, tais como: avaliação, monitoramento, acompanhamento, indicadores e suas propriedades e tipos de avaliação. Esta seção presta-se não apenas a contribuir para a uniformização conceitual, mas também a apresentar os tipos de avaliação priorizados pelo Departamento, os quais se associam às funções que lhe cabe cumprir e aos seus métodos e parcerias.

As seções seguintes apresentam princípios, requisitos, funções da avaliação e critérios para decisão sobre como avaliar, além dos métodos correspondentes. Cabe destacar mais uma vez que não se trata apenas de fazer referência à literatura de avaliação, mas também de revelar opções metodológicas e procedimentais feitas pelo DAIE, à luz das possibilidades existentes e das boas práticas já construídas pela

comunidade de avaliação, formada por acadêmicos, gestores e instituições especializadas.

Por fim, tem-se uma seção de comentários finais, que reafirma os propósitos deste texto e indica como se espera que ele venha a contribuir para o trabalho do DAIE nesta nova etapa.

1. Avaliação e Monitoramento como parte do ciclo de gestão: a experiência da administração pública federal e do Departamento de Avaliação e Informações Educacionais

1.1 – Avaliação como função gerencial orientada para a busca de resultados

A avaliação e o monitoramento de políticas públicas como etapa integrante do ciclo de gestão pública foi introduzida na administração pública federal apenas recentemente, como parte de um processo de mudança no cenário das políticas públicas no Brasil, caracterizado principalmente pela reforma administrativa e sua ênfase na busca por resultados. A importância da avaliação e do monitoramento se maximiza num contexto de limitação de recursos e de foco nos resultados, mais do que nos processos. No Brasil, a avaliação e o monitoramento também se fazem mais relevantes como parte integrante de uma transformação no modelo de gestão das políticas públicas, especialmente das políticas sociais. O principal elemento desta transformação diz respeito ao aumento da racionalidade nos processos decisório e gerencial de modo a tornar as políticas sociais mais efetivas, eficazes e eficientes. A provisão de racionalidade se coloca como um desafio ainda maior na medida em que a redemocratização provoca a afluência de uma variedade de atores e de demandas ao espaço público.

Nas décadas de 80 e 90, manifesta-se no Brasil uma crise fiscal que leva ao questionamento das políticas sociais e do papel do Estado enquanto agente econômico ativo e provedor de serviços sociais universais. Ao mesmo tempo, uma série de transformações econômicas e sociais (o envelhecimento populacional, o surgimento do desemprego estrutural, novos modelos de família, novas ocupações, redução dos níveis de sindicalização, aumento dos fluxos internacionais de comércio e de intercâmbio cultural, globalização financeira etc) impõem novas demandas ao Estado enquanto este ainda luta por resolver dilemas antigos, notadamente os altos níveis de desigualdade social e pobreza e, no campo da educação, a falta de acesso da maior parte da população à instrução formal. Neste contexto, o modelo híbrido de administração pública brasileiro, em que elementos de clientelismo e de patrimonialismo subsistem ao lado de um modelo racional-burocrático inacabado, parece claramente inadequado para um Estado capaz de atuar de forma mais efetiva. As demandas sociais impõem a implantação de um modelo de administração pública com enfoque nos resultados.

O conceito de *governança*, que adquire proeminência nas discussões sobre

reforma do Estado, envolve três dimensões fundamentais, especificadas por Diniz (apud Santos Júnior 2001, p. 55): “capacidade de coordenação do Estado entre as distintas políticas e os diferentes interesses em jogo, capacidade de comando e de direção do Estado e capacidade de implementação”. A promoção de uma boa governança requer que o Estado tenha capacidade para coordenar todos os atores políticos envolvidos no processo decisório de políticas públicas, que se tornam mais numerosos num contexto democrático, e que estas alcancem os objetivos propostos ao menor custo possível. Neste sentido, a avaliação e o monitoramento podem ser compreendidos como um instrumento de promoção da governança democrática, que contribui para tornar mais transparentes as relações entre Estado e sociedade e para assegurar que o Estado seja capaz de responder adequadamente às demandas sociais, implementando intervenções ajustadas à resolução dos problemas diagnosticados.

A avaliação é uma ferramenta importante, portanto, para orientar que políticas públicas, dentre as várias possibilidades existentes, se deve implementar. Políticas públicas devem ser entendidas aqui como intervenções públicas destinadas a cumprir determinados objetivos e atuar sobre dimensões sociais amplas, orientadas por preceitos políticos e técnicos. Utilizando-se a terminologia do Plano Plurianual, tem-se que as políticas são compostas por programas - instrumentos de organização da ação governamental com vistas ao enfrentamento de um problema e suas causas. Por sua vez, as ações constituem atividades ou projetos, que possuem escopo menor e limitado temporalmente e que se articulam para alcançar o objetivo do programa.

A concorrência entre políticas públicas impõe ao tomador de decisão a necessidade de adotar critérios técnicos, ademais de políticos, para selecionar a alternativa de solução que seja capaz de gerar a maior gama de resultados ao menor custo possível. Avaliações *ex-ante*, como será detalhado mais adiante, destinam-se a aferir os impactos potenciais, *vis-a-vis* os custos, de uma ação governamental, em geral quando esta ação ainda não começou a ser implementada. No entanto, a seleção da melhor alternativa requer que as ações públicas (políticas, programas ou projetos) sejam avaliáveis, isto é, sejam desenhadas desde o início para permitir sua avaliação. Um dos requisitos para isso é que seja possível identificar a “teoria do programa” que orienta a intervenção pública, isto é, a explicitação “da cadeia de premissas que explicam como as atividades de implementação do programa levam, passo a passo, aos resultados desejados” (Weiss, 1988, p. 265). Supõe-se assim que todo programa ou projeto público esteja orientado por uma teoria que relaciona meios (bens ou serviços prestados no

âmbito do programa) a fins (resultados que se espera atingir através daqueles meios). A identificação da teoria do programa facilitará a execução de avaliações que poderão responder por que o programa foi capaz ou incapaz de atingir determinados resultados e que fatores alavancaram ou reduziram o sucesso do programa.

Para que esta teoria seja válida, é preciso que o gestor do programa esteja pautado em informações confiáveis a respeito da natureza do problema que se quer combater, bem como de suas causas. É preciso, portanto, elaborar um "diagnóstico social" (Paes de Barros e Carvalho, 2006, p. 3), o qual, idealmente, deveria estar baseado num sistema de monitoramento de resultados sociais, ou seja, de indicadores gerais relativos à dimensão social sobre a qual se trabalha: educação, saúde, trabalho etc. A partir da análise destes indicadores, torna-se possível estabelecer, com mais precisão, relações de causa e efeito, que são o elemento principal de uma teoria do programa. Ao desenhar novas políticas públicas, é importante também buscar informações a respeito de ações semelhantes existentes em outros países e suas avaliações, que informarão o gestor a respeito de fatores associados ao desempenho daquelas ações.

1.2 – Avaliação e Monitoramento no contexto da administração pública federal: o PPA e outros exemplos

No âmbito da administração pública brasileira, a necessidade de avaliar torna-se imperativa dada a existência de um descompasso entre demandas e recursos, ou seja, recursos escassos e necessidades ilimitadas. Além disso, tem-se proclamado a exigência de aumentar a qualidade (em detrimento da quantidade) dos gastos públicos na área social. Também neste sentido a avaliação funciona como um poderoso subsídio para o processo decisório acerca de seleção e manutenção de políticas. No embate político que envolve a inserção de alternativas de política na agenda e seleção das políticas a serem adotadas, os argumentos técnicos baseados em avaliações podem ser de grande valia para os atores envolvidos.

Na realidade, o reconhecimento das funções exercidas pelos sistemas de avaliação e monitoramento e de sua importância para uma gestão orientada pela busca de resultados já está inscrito na cultura de planejamento da administração pública federal por meio do Plano Plurianual. No âmbito do PPA, a avaliação é descrita como necessária por "contribuir para o alcance dos resultados dos programas, por meio do aperfeiçoamento da gestão de programas e da alocação de recursos no PPA e nos

orçamentos da União, de modo a aprimorar a qualidade do gasto público”[1]. O modelo de gestão do PPA contém um Sistema de Monitoramento e Avaliação, integrado por uma Comissão de Monitoramento e Avaliação, de caráter interministerial, e por Unidades de Monitoramento e Avaliação, existentes em todos os ministérios e secretarias especiais, que procuram assegurar que estas duas funções sejam executadas de forma integrada. Desta forma, os gerentes de programas estão já submetidos à exigência de preparação de avaliações anuais dos programas sob sua responsabilidade que abrangem tanto resultados, quanto concepção e implementação dos programas. O PPA também prescreve elementos de avaliações ex-ante na medida em que requer a formulação de índices de referência relativos à situação anterior ao início da execução do programa e ao longo do período de vigência do plano[2].

A exemplo de países como Estados Unidos, Austrália, Canadá e França - onde os órgãos governamentais que fazem auditoria estão sendo solicitados a realizar avaliações -, no Brasil, o Tribunal de Contas da União vem realizando avaliações pontuais de alguns programas governamentais, o que pode certamente contribuir para a disseminação de uma cultura de avaliação na administração pública. De forma análoga, a Controladoria Geral da União promove auditorias de desempenho, que implicam avaliar em que medida as políticas, programas e instituições governamentais estão sendo bem gerenciadas. Apesar de ter por função precípua a realização de auditorias e fiscalizações, a CGU também avalia o cumprimento das metas previstas do PPA e a execução das ações de governo e dos orçamentos da União. Paulatinamente, estes órgãos têm ampliado o seu foco, procurando averigüar também a eficiência e a eficácia, além da legalidade e da economicidade da gestão pública. Há uma complementaridade entre as avaliações promovidas pelo DAIE e pelas instituições de controle interno e externo na medida em que estas contribuem mutuamente para a viabilização de suas funções, isto é, umas servem como subsídio às outras. Todos estas avaliações também servem ao propósito comum de fortalecer o controle social e a *accountability*[3] da administração pública.

---

[1] Manual de Avaliação do Plano Plurianual 2004-2007. Brasília, 2003, p. 3.

[2] Todavia, ainda se faz presente a necessidade de que este modelo prescrito pelo Ministério do Planejamento seja plenamente desenvolvido, ou seja, que todas as instituições integrantes do Sistema de Monitoramento e Avaliação estejam em pleno funcionamento.

[3] Segundo Pó e Abrúcio, (2006, p. 686), “*accountability”* é um conceito amplo, diz respeito à “existência de condições para os cidadãos participarem da definição e avaliação das políticas públicas, premiando ou punindo os responsáveis (Clad, 2000). Nessas condições deve constar a disponibilidade de informações sobre a atuação dos governos e seus resultados, bem como a existência de instituições que permitam contestar as ações do poder público”.

O Ministério da Educação tem grande experiência e tradição em avaliação de políticas educacionais, o que tem sido priorizado e aprimorado nesta gestão. Destaca-se o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), autarquia vinculada ao MEC, encarregada de promover estudos, pesquisas e avaliações sobre o Sistema Educacional Brasileiro com o objetivo de subsidiar a formulação e implementação de políticas públicas para a área educacional. O INEP promove avaliações sistêmicas de resultados de todos os níveis de ensino, como o Sistema Nacional de Avaliação da Educação Básica (Saeb) e o Sistema Nacional de Avaliação da Educação Superior, além de levantamentos censitários sobre o sistema educacional, como o Censo Escolar e outros. Ao lado disso, o INEP também promove avaliações cognitivas individuais, como o Exame Nacional do Ensino Médio (Enem) e o Prova Brasil, realizado pela primeira vez em 2005. Esta última, em conjunto com o SAEB, constituem as fontes de dados necessárias para composição do Ideb – Índice de Desenvolvimento da Educação Básica, calculado para todos os municípios e escolas a fim de avaliar o aprendizado dos alunos, o qual constitui um dos componentes principais do Plano de Desenvolvimento da Educação (PDE), lançado pela atual gestão. A avaliação possui, portanto, um lugar fundamental neste Ministério. Ademais das avaliações dos resultados finais alcançados pelo sistema educacional federal, estadual e municipal, tem-se também priorizado a realização de avaliações com foco no desempenho e na gestão de políticas e programas específicos, visando corrigir rumos e propor redesenhos.

Neste sentido, existem, na administração pública federal, ao menos três tipos de localização institucional das áreas responsáveis por avaliação e monitoramento de políticas e programas: (i) as Unidades de Monitoramento e Avaliação (UMAs), situadas em geral nas Subsecretarias de Planejamento, Orçamento e Administração de cada órgão; (ii) a Secretaria de Avaliação e Gestão da Informação do Ministério do Desenvolvimento Social e Combate à Fome; e (iii) o Departamento de Avaliação e Informações Educacionais da SECAD.

As chamadas UMAs, já citadas anteriormente, integram o Sistema de Avaliação do Plano Plurianual e têm por função apoiar cada órgão na avaliação e monitoramento dos programas e na elaboração de planos gerenciais. São elas as responsáveis pela consolidação das Avaliações anuais do Plano Plurianual e interlocução junto ao Ministério do Planejamento.

A Secretaria de Avaliação e Gestão da Informação tem a função de promover a

avaliação e o monitoramento das políticas e programas de desenvolvimento social do Ministério do Desenvolvimento Social e Combate à Fome. As peculiaridades da SAGI em relação aos demais modelos existentes na administração pública federal estão relacionadas aos fatos de constituir uma secretaria exclusiva para a área de monitoramento e avaliação e de estar localizada horizontalmente em relação às secretarias finalísticas. Tendo sido criada recentemente (2004), a SAGI tem logrado constituir um sistema de avaliação e monitoramento para todo o Ministério.

O Departamento de Avaliação e Informações Educacionais (DAIE) do MEC ocupa outro lócus institucional, ou seja, está situado dentro de uma secretaria também criada recentemente (julho de 2004). Sua principal função consiste em acompanhar, monitorar e avaliar os programas e ações implementadas pela SECAD, além de realizar estudos que subsidiem a construção de indicadores para os programas e a formulação de novas políticas educacionais. Por ter um escopo menor de atuação em comparação com a SAGI, dado que sua missão restringe-se à SECAD, o DAIE tem a oportunidade de executar suas funções com maior detalhamento e profundidade visto que é menor o número de programas e ações que cabe a este departamento monitorar e avaliar. Além disso, o fato de constituir um departamento, em vez de uma secretaria, possibilita à equipe técnica de monitoramento e avaliação uma aproximação maior e mais constante com os departamentos finalísticos. Esta proximidade resulta bastante vantajosa para a execução da função que cabe ao departamento dado que potencializa a execução da função precípua da avaliação, que é a de retroalimentar a gestão em todos os seus momentos. Pode-se afirmar ainda que o pertencimento à mesma secretaria possibilita um maior compartilhamento de interesses e objetivos entre a equipe do DAIE e as demais equipes, o que se reflete em relações profissionais pautadas por maior confiança e cooperação. Por outro lado, a constituição como departamento provê à área de avaliação de políticas e programas de educação continuada, alfabetização e diversidade menor projeção interna e externamente ao MEC. O orçamento disponível para o departamento também tende a ser menor, comparativamente ao que teria uma secretaria.

Para fechar esta seção, deve-se reafirmar o que já foi dito explicita ou implicitamente, a saber: que a avaliação está presente em todo o ciclo de gestão, desde a inserção do problema na agenda pública, passando pelas etapas seguintes de desenho e formulação de alternativas de ação, implementação da ação e, por último, a aferição dos resultados intermediários e finais e o uso dos resultados da avaliação para a tomada de decisão a respeito do futuro do projeto. Idealmente, a avaliação deveria gerar subsídios

para todas estas etapas, ao contrário do que apregoa a versão tradicional, para a qual a avaliação se faz apenas ao final, como um rito de fechamento ou balanço de um projeto, no intuito de averigüar seus sucessos e fracassos. Essa não é a visão adotada pelo DAIE. Assim como as políticas públicas não se desenvolvem em etapas seqüenciais rígidas - visto que a formulação muitas vezes ocorre concomitantemente à implementação -, o DAIE advoga que a avaliação não deve estar restrita ao final do processo; ao contrário, deve subsidiá-lo durante todo o percurso, em estreita colaboração de técnicos em avaliação com gestores de áreas finalísticas. O DAIE considera ainda que as especificidades da SECAD, a qual procura atuar sobre os padrões de desigualdade da sociedade brasileira, bem como sobre temas relacionados à diversidade ainda não consensuais entre gestores e a sociedade em geral (tais como orientação sexual e identidade de gênero, relações étnico-raciais e outros), aumentam a imprescindibilidade de que seus programas e ações sejam executados com eficiência, eficácia e efetividade e de que seus resultados sejam amplamente publicados. A avaliação, o monitoramento e o acompanhamento têm uma contribuição fundamental para o alcance desses objetivos.

2.Definições conceituais básicas para a área de Monitoramento e Avaliação

2.1 Avaliação e classificações

A definição de conceitos básicos contribui para a sua uniformização e facilita a comunicação entre o DAIE e as demais unidades da SECAD e do MEC. O primeiro conceito básico que se pretende definir é o de avaliação, entendida aqui como um processo de coleta e análise sistemáticas de informações sobre insumos, processos e resultados de uma ação governamental (política, programa ou projeto), com base em critérios de eficiência, eficácia e efetividade, de forma a gerar recomendações para aperfeiçoar a gestão e a qualidade do gasto público. Embora esta não seja a única definição possível, ela é ampla o bastante para compreender todos os tipos de avaliação que podem e devem ser executados para subsidiar os gestores e que serão detalhados a seguir.

Seguindo a classificação proposta pelo Manual de Avaliação do PPA, as avaliações podem ser classificadas, quanto à temporalidade, em dois tipos:

- Avaliações *ex-ante*: destinam-se a aferir os impactos potenciais de uma ação governamental, em geral quando esta ação ainda está na fase de desenho, ou seja, não começou a ser implementada. É preciso que a avaliação *ex-ante* seja pensada desde a fase de formulação da ação para que ela possa subsidiar a tomada de decisão a respeito

de qual seria a melhor alternativa para solução do problema diagnosticado ao qual se dirige a ação. Para isso, requer a projeção de custos e impactos potenciais das alternativas de ação.

- Avaliações *ex-post* são realizadas quando a ação já existe, em qualquer fase, mesmo que seja no piloto. Em geral, utiliza dados gerados pela própria ação que se quer avaliar, oriundos do monitoramento, embora possa ser subsidiada por dados secundários de diversas fontes (Paes de Barros e Carvalho, 2006, p. 15). Este é o tipo de avaliação mais realizado no âmbito dos órgãos de governo, incluindo a SECAD, e pode abranger uma variedade de objetos (recursos, processos e resultados) e métodos. Embora tenda-se a associar avaliações ex-post com avaliações de impacto, não existe esta restrição.

Quanto ao objeto, pode-se classificar as avaliações nos seguintes tipos:

- Avaliações de processos: procuram aferir a qualidade da gestão dos programas, utiliza uma variedade de métodos e indicadores, tanto quantitativos quanto qualitativos, que procuram medir o grau de capacidade institucional e a eficiência[4] e eficácia[5] da gestão, em geral correlacionando estes aspectos com os resultados gerados pela ação. A avaliação de processos tem como objeto a implementação das ações, qualquer que seja o estágio desta.

- Avaliações de resultados: procuram mensurar os resultados intermediários (produtos) e/ou finais (impactos), obtidos pela ação, dados os seus objetivos. No entanto, uma ação pode também gerar resultados não relacionados aos seus objetivos, ou seja, pode influenciar outras dimensões sociais sobre as quais a ação não interveio diretamente. Esses resultados, que podem ser antecipados ou não, são chamados de efeitos (Draibe, 2001, p. 21). Assim como as avaliações de processos, utilizam uma diversidade de métodos e concentram-se sobre a eficácia e a efetividade[6] das ações.

- Avaliações de impacto: destinam-se a aferir os resultados finais atingidos pela ação sobre determinadas dimensões que a ação objetiva afetar. Constituem um sub-

---

[4] Eficiência, segundo a terminologia proposta pelo Plano Plurianual (Decreto nº 5.333), é entendida como a medida da relação entre os recursos efetivamente utilizados para a realização de uma meta para um projeto, atividade ou programa frente a padrões estabelecidos.

[5] Eficácia diz respeito à medida do grau de atingimento das metas fixadas para um determinado projeto, atividade ou programa em relação ao previsto. A eficácia refere-se, portanto, à consecução dos resultados intermediários ou de curto prazo (produtos) obtidos pela ação.

[6] Efetividade refere-se à medida do grau de atingimento dos objetivos que orientaram a constituição de uma determinada ação.

tipo específico de avaliação de resultados e procuram isolar o impacto da ação sobre as variáveis que se pretende modificar para medir a efetividade desta ação. Por isso, requerem a utilização de métodos específicos de alto rigor científico. Devem ser realizadas passado algum tempo do início da implementação da ação, quando se supõe que ela já tenha sido capaz de gerar impactos de médio e longo prazo sobre os beneficiários das ações.

É importante destacar que estes tipos de avaliações podem ser combinados, ou seja, uma única avaliação pode aferir aspectos relacionados a processos e a resultados.

Quanto à execução, as avaliações podem ser classificadas em internas, externas e mistas.

- Avaliações internas: são as avaliações realizadas pelas equipes técnicas das áreas finalísticas, com apoio do DAIE. São avaliações de pequena escala, que utilizam dados produzidos pela própria gestão da ação ou dados secundários. Em geral, são avaliações pouco dispendiosas e tendem a produzir uma alta retroalimentação da gestão visto que são conduzidas pelos próprios gestores.
- Avaliações externas: contratadas pela SECAD e executadas por instituições parceiras. Avaliações externas são supervisionadas pelo DAIE, que institui o desenho básico da avaliação no plano de avaliação, incluindo procedimentos metodológicos, e encarrega-se de colaborar com as instituições parceiras na definição dos instrumentos de coleta de dados e análise dos resultados. O DAIE também checa a qualidade dos produtos apresentados pelas instituições parceiras, assegurando que estes obedeçam a padrões de excelência.
- Avaliações mistas: internas e externas. São avaliações realizadas em conjunto com instituições externas. Neste caso, as etapas que compõem as avaliações são divididas ou compartilhadas entre a SECAD e a instituição contratada. Este tipo de avaliação tem o potencial de promover o aprendizado institucional e a capacitação da equipe, além de também assegurar o aporte técnico de instituições especializadas.

Quanto ao propósito a que se destinam, as avaliações podem ainda ser classificadas em:

- Avaliações formativas: “são desenhadas para assistir aqueles que estão desenvolvendo ou gerenciando as ações” (Weiss, 1998, p. 31). Devem, preferencialmente, ser realizadas nos estágios iniciais de implementação das ações para que possam gerar recomendações úteis para a obtenção de mais/melhores resultados.

Seu objeto são os processos de implementação.

- Avaliações somativas: sua ênfase recai sobre os resultados da ação e geralmente servem ao propósito de gerar recomendações a respeito do aperfeiçoamento da ação ou de sua extensão, redução ou demais alterações substantivas. Em último caso, este tipo de avaliação pode subsidiar uma decisão de extinguir ou não uma determinada ação, o que ocorre muito raramente.

O DAIE entende a avaliação e o monitoramento como uma etapa imprescindível da gestão de seus programas. Para este departamento, a atuação conjunta com os gestores das áreas finalísticas da secretaria reveste-se de importância na medida em que potencializa a utilidade das avaliações, isto é, a incorporação de seus resultados na gestão dos programas, além de aprimorar o desenho das avaliações e a relevância das questões propostas por elas. Portanto, o tipo de avaliação privilegiado pelo DAIE é a formativa, aquela que pode efetivamente gerar resultados úteis à gestão das ações promovidas pela Secretaria. Esta escolha preferencial por avaliações formativas, porém, não exclui a possibilidade de realizar avaliações somativas, as quais também produzem resultados úteis aos gestores.

2.2. Monitoramento e acompanhamento

Monitoramento e avaliação são processos de gestão complementares, dado que ambos servem ao propósito de aperfeiçoar a formulação e a implementação de políticas e programas públicos. O monitoramento consiste na análise contínua e sistemática dos programas, verificando a realização dos processos, a alocação dos recursos e a geração dos produtos, com base em indicadores que aferem a implementação em relação às etapas, finanças e metas previamente fixadas. Por meio dele, o gestor adquire a capacidade de intervir de forma tempestiva para correção de rumo dos programas, atuando diretamente sobre as fragilidades apontadas pelos resultados dos indicadores. Pode ser realizado a distância ou in loco, utilizando-se instrumentos e procedimentos de pesquisa empírica para coleta de dados que possibilitarão a alimentação dos indicadores relevantes.

O monitoramento se distingue da avaliação na medida em que produz e atualiza continuamente e sistematicamente informações sobre a implementação das ações, registrando aspectos qualitativos e quantitativos sobre os processos, recursos e produtos a partir de informações prestadas pelos coordenadores de ação para orientar a intervenção do programa. Não obstante, o monitoramento não faz análises explicativas,

com a utilização de métodos científicos, sobre a correlação entre processos e resultados, tal como faz a avaliação.

O acompanhamento é um processo que auxilia a realização do monitoramento, subsidiando os gestores na definição de critérios e parâmetros, bem como na preparação de sistemas de informação gerenciais. No âmbito da SECAD, o monitoramento é uma atividade que cabe aos gestores e gerentes das coordenações finalísticas desempenhar, enquanto que o acompanhamento cabe ao DAIE, que possui uma coordenação especializada para analisar e consolidar os dados do monitoramento enviados pelas coordenações finalísticas e levantados nas observações *in loco* para elaborar relatórios periódicos de gestão.

Por sua vez, o conceito de indicador pode ser definido, segundo Jannuzzi (2003, p. 15), como "uma medida, em geral quantitativa, dotada de significado social substantivo, usado para substituir, quantificar ou operacionalizar um conceito social abstrato, de interesse teórico (para pesquisa acadêmica) ou programático (para formulação de políticas). É um recurso metodológico, empiricamente referido, que informa algo sobre um aspecto da realidade social ou sobre mudanças que estão se processando na mesma". Os tipos de indicadores propostos pela SECAD relacionam-se aos interesses e necessidades dos gestores de políticas públicas e parceiros, quais sejam:

- Indicadores de Estrutura ou Insumo: medem a quantidade de recursos de diversos tipos que são aportados ao programa/ação para viabilizar sua implementação. Podem ser classificados em: humanos, físicos e financeiros;
- Indicadores de Processo: referem-se aos procedimentos e modos de conversão e utilização dos recursos aportados para gerir o programa na forma de processos necessários à implementação das ações;
- Indicadores de resultados: medem o grau em que os objetivos intermediários ou finais da ação têm sido atingidos. Podem ser divididos em três subtipos:
    - Indicadores de produto referem-se aos resultados intermediários ou de curto prazo (como número de beneficiários que concluíram cursos de capacitação) obtidos pela ação.
    - Indicadores de impacto são relativos aos resultados de mais longo prazo (como número de famílias atendidas que superaram a condição de vulnerabilidade). Avaliações de impacto, destinadas a isolar os impactos de programas, utilizam

indicadores de impacto como base da investigação. Os indicadores de impacto são mais dificilmente calculáveis devido às dificuldades de se atribuírem os resultados à intervenção promovida pelo programa.

- Indicadores de efeitos medem os impactos indiretos, antecipados ou não, provocados pelo programa. Segundo a classificação de Draibe (2001, p. 25), os efeitos podem ser divididos em sociais, quando incidem sobre "os agentes implementadores, a comunidade local e grupos particulares de interessados na sua execução" e institucionais, que ocorrem quando a ação afeta as instituições governamentais e não-governamentais envolvidas na sua implementação.

Há uma série de dificuldades envolvidas em estabelecer relações de causalidade para composição de indicadores de impactos e efeitos. Existem determinados recursos metodológicos e analíticos que podem assegurar um alto grau de confiabilidade na correlação entre variáveis dependentes (aspectos observáveis do fenômeno que se pretende explicar) e independentes ou explicativas, que afetam a variável dependente e provocam nela uma mudança. Ainda assim, a aplicação destes recursos exige uma série de requisitos em termos de linha de base, composição da amostra, coleta de dados e uso de hipóteses que tornam as avaliações de impacto bastante complexas.

Em geral, os gestores têm mais governabilidade sobre produtos e, por isso, os indicadores de produto são mais facilmente calculáveis. No entanto, estes indicadores em geral não informam sobre a capacidade da ação em alcançar seus objetivos finais, para os quais foram concebidas.

2.2.1 Propriedades dos indicadores

Os indicadores possuem determinadas propriedades cuja análise revela o grau de acerto na sua elaboração e utilização. A validade corresponde ao grau de proximidade entre o conceito (por exemplo, infra-estrutura educacional) e a medida (por exemplo, número de escolas). Ao avaliar a validade de um indicador, deve-se testar em que medida ele é capaz de refletir com exatidão aquilo que se quer expressar por meio do conceito. A confiabilidade diz respeito à qualidade dos dados utilizados para composição do indicador. Pode-se questionar a confiabilidade de um indicador se os dados utilizados são incompletos ou se a fonte é inadequada. Sensibilidade é a propriedade que informa sobre a capacidade do indicador de refletir mudanças significativas no fenômeno ao qual ele se refere. Supõe-se que indicadores sensíveis (por exemplo, taxa de analfabetismo) mudem à medida que o fenômeno se altere

(número de alunos aprovados em cursos de alfabetização para jovens e adultos). Outras propriedades relevantes são a periodicidade, ou seja, a freqüência e o custo com os quais o indicador pode ser atualizado e a inteligibilidade, isto é, a facilidade com que ele pode ser compreendido e o quão transparente é a metodologia de sua composição.

Ao selecionar indicadores para avaliação e monitoramento, é preciso checar como eles se comportam em relação a estas propriedades. Do contrário, pode-se concluir que determinada ação não está produzindo resultados quando, na verdade, pode-se estar utilizando indicadores de resultados inválidos (como taxa de desemprego para programas de qualificação de mão de obra) ou muito ambiciosos comparativamente ao escopo das ações promovidas.

3. Princípios e funções da avaliação

3.1 Princípios

Toda avaliação deve ser regida por determinados princípios, tais como: isenção e independência em relação ao objeto que está sendo avaliado de modo a obter resultados fidedignos; articulação com os gestores das ações avaliadas e outros atores de forma a torná-los parceiros na avaliação; respeito aos atores envolvidos na avaliação (pessoas a serem entrevistadas) e compromisso com a confidencialidade das informações recebidas; publicação dos resultados (visto que se trata de programas sociais públicos e que as informações sobre eles interessam ao público em geral) e retorno sobre eles para os atores que colaboraram na avaliação.

3.2 - Função da Avaliação

Os princípios devem funcionar como balizadores da ação dos avaliadores para assegurar que as funções da avaliação serão exercidas. Trata-se aqui de quatro funções principais.

A primeira delas é a de **aprendizado** para aqueles que estão executando uma determinada ação. Por meio de avaliações *ex-post*, os gestores adquirem a informação necessária para decidir a respeito da viabilidade, do custo-efetividade das alternativas disponíveis, bem como a respeito da qualidade dos processos de implementação e dos resultados alcançados. Como já ressaltado anteriormente, a avaliação provê informação que possibilita ao gestor apreender o funcionamento e as conseqüências de suas decisões num determinado momento, podendo passar ao seguinte com maior segurança. Como afirma Rua (s.d., p. 2), a avaliação é “decisiva para o processo de aprendizagem

institucional" pois, por meio dela, as instituições adquirem ganhos de racionalidade no processo decisório e de conhecimento válidos não apenas para a ação avaliada, mas também para diversas outras que a ela se assemelham.

A segunda função é de **transparência,** pois a publicação dos resultados de avaliações oferece à sociedade uma fonte preciosa de informação a respeito de como os recursos públicos estão sendo gastos e com que eficiência, eficácia e efetividade. As avaliações provêem à sociedade subsídios importantes também ao exercício do controle social das políticas públicas. Além disso, as avaliações em geral possibilitam que o público-alvo das ações se manifeste, por meio de entrevistas, a respeito dos resultados produzidos por elas em sua qualidade de vida, o que as torna um elemento importante para a democratização do processo de gestão. O incremento da transparência tende a refletir-se também em aumento de legitimidade social e política das ações.

A terceira é a função estratégica na **escolha** de prioridades e no investimento de recursos. Avaliações *ex-ante*, como já explicado, servem ao propósito de proporcionar ao gestor os instrumentos necessários para escolher uma alternativa de ação dentre várias possíveis. Além da escolha entre ações, a avaliação *ex-post* também auxilia o gestor no processo de fazer escolhas a respeito da ação sob sua responsabilidade que já está em execução. Escolhas relativas a alocação de recursos, como necessidade de aportar maiores recursos humanos ou financeiros para programas que apresentam problemas de implementação, são típicas decisões que podem ser subsidiadas por avaliações. Avaliações também subsidiam escolhas ou decisões que são feitas por instâncias decisórias superiores ou pelos próprios gestores em conjunto com outras instâncias. Por exemplo, o relacionamento dos gestores com instituições externas, como Secretaria de Planejamento e Investimentos Estratégicos (SPI) e o Instituto de Pesquisa Econômica Aplicada, do Ministério do Planejamento, que podem fazer escolhas ou tomar decisões capazes de intervir sobre o futuro das ações públicas, também pode ser facilitado quando se usam avaliações como argumentos técnicos poderosos num embate.

Desta forma, uma avaliação deve ter a preocupação de servir às necessidades informacionais dos gestores, gerando conhecimento relevante para o processo decisório. É neste aspecto que ela se distingue de pesquisas científicas básicas, com as quais compartilha os mesmos princípios metodológicos e das quais diverge em relação aos objetivos.

A quarta função é a de proporcionar a informação necessária para correção de

rumos e **redesenho** dos programas. Para facilitar o processo de incorporação dos resultados das avaliações, é preciso que elas disponham de recomendações precisas, baseadas na identificação dos determinantes do sucesso e do fracasso no alcance dos resultados esperados. Embora se costume difundir a possibilidade de extinguir ações a partir de avaliações, em geral isso não acontece. É mais comum que elas sofram alterações de menor escopo, isto é, um redesenho da ação. O redesenho não precisa ocorrer apenas ao final da avaliação de resultados; ele pode acontecer em etapas anteriores de modo a possibilitar a correção a tempo de fragilidades e aspectos negligenciados que deveriam receber mais atenção por parte dos gestores de forma a potencializar a efetividade das intervenções.

3.3 - Requisitos da avaliação

Os requisitos postos abaixo (monitoramento, registros administrativos e bases de dados consistentes) provêem robustez às funções da avaliação. Se cumpridos, possibilitarão avaliações de melhor qualidade, que poderão gerar recomendações úteis.

- Avaliabilidade: refere-se à possibilidade de que a ação seja avaliada e depende de uma série de fatores, a saber: que a ação a ser avaliada tenha um modelo lógico[7], relacionando causa e efeito e meios a fins; que as perguntas avaliativas sejam formuladas e os procedimentos metodológicos de investigação sejam estabelecidos; que os dados necessários sejam disponíveis ou passíveis de serem coletados; que haja recursos para realizar a avaliação.
- Monitoramento. Idealmente, a avaliação deveria utilizar as informações geradas pelo monitoramento para responder a questões mais sofisticadas, que demandam a execução de estudos em profundidade acerca de determinados aspectos da ação, como por exemplo, o estabelecimento de relações causais entre variáveis dependentes, supostamente afetadas pelo programa, e variáveis independentes. Ao monitorar uma ação, deve-se ter em vista que o monitoramento deve servir à viabilização de futuras avaliações: por isso, os indicadores monitorados devem ser válidos, sensíveis, confiáveis, inteligíveis e periodicamente atualizados. Além disso, deve-se pensar em indicadores de estrutura, processos e resultados para viabilizar avaliações de diversos tipos.
- Registros administrativos. A produção constante e manutenção de

[7] Ver Box na página X.

registros administrativos é imprescindível para a avaliação, o monitoramento e o acompanhamento visto que tais registros constituem um insumo chave para o acompanhamento dos indicadores. A realização de pesquisas longitudinais e de estudos que pretendem aferir resultados por meio de comparações antes-depois, por exemplo, se beneficiam da manutenção de registros administrativos a respeito dos processos de implementação dos programas e dos cadastros de beneficiários. Assim como no caso do monitoramento, deve haver também a preocupação dos gestores em assegurar que os dados sejam registrados de forma objetiva e o mais completa possível (por exemplo, cadastros de beneficiários devem conter, minimamente, dados sobre nome, endereço, telefone, CPF, escolaridade, data de nascimento e outros dados considerados relevantes para o monitoramento e a avaliação de cada ação específica).

- Base de dados consistente. Em geral, a acuidade das avaliações requer a existência de linhas de base, que consistem em bases de dados que contenham os indicadores relevantes referentes ao período anterior ao início da implementação da ação. Em avaliações de impacto, a comparação dos dados relativos ao período pós-intervenção com dados anteriores contribui para que se possa isolar o impacto do programa com certa margem de segurança. Desta forma, pode-se averigüar em que medida a dimensão social foi afetada pelo programa. Independentemente do tipo de avaliação a ser feita, no entanto, é necessário dispor de bases de dados primários e secundários consistentes, ou seja, é preciso proceder à crítica dos dados, rechecagem dos questionários em caso de aparentes desvios e realização de testes de consistência.

4. Decidindo como e quando avaliar: critérios e métodos

4.1 – Critérios e procedimentos do DAIE

O DAIE, em conjunto com os demais departamentos integrantes da secretaria, encarrega-se de identificar as necessidades de avaliação junto aos gestores dos programas e projetos, desenhar e executar ou, em alguns casos, supervisionar a execução de avaliações contratadas externamente.

Ao decidir que ações avaliar, o DAIE utiliza diversos critérios, dos quais os mais importantes são:

a) relevância da ação para a política pública educacional de promoção da diversidade, indicada por fatores como o grau de inovação da ação (introdução de novos temas na agenda), orçamento, grau de cobertura, quantidade de parcerias envolvidas e outros atributos;

b) grau de avaliabilidade e necessidades de informação acerca da ação que seriam supridas por meio da avaliação.

A decisão sobre como avaliar pauta-se em critérios como:

c) aspectos da ação que precisam ser avaliados e métodos que podem ser empregados;

d) tempo, orçamento e recursos humanos disponíveis para avaliação.

Todas as ações podem ser avaliadas de diversas maneiras. Mesmo que a ação não tenha ainda começado a produzir resultados, poderá ser submetida a uma avaliação formativa que proverá informações aos gestores para a transição aos próximos estágios de gestão. Trata-se, portanto, de elaborar um plano de avaliação, contendo um diagnóstico da ação a ser avaliada, os tipos possíveis de avaliação a serem feitas, bem como os requisitos de informação, os procedimentos metodológicos e informações sobre custo e tempo de execução associados a cada uma delas. O plano oferece subsídios sobre como selecionar as avaliações que devem ser priorizadas de acordo com o estágio em que se encontra a ação e como a avaliação pode favorecer o seu aperfeiçoamento.

A decisão sobre o lócus de execução da avaliação depende da função a que deve servir a avaliação e dos princípios que a regem. Em geral, avaliações internas são realizadas quando o escopo da avaliação é pequeno, com a utilização de dados secundários ou produzidos pela própria gestão e quando há disponibilidade de recursos humanos para sua execução. Avaliações internas são mais propícias ao aprendizado, porém deve haver isenção e independência para sua realização. Avaliações externas são realizadas quando há necessidade de coleta de dados in loco para grandes amostras. O volume de dados a serem analisados em geral inviabiliza a realização deste tipo de avaliação por parte da própria secretaria. Além disso, é desejável que algumas avaliações sejam contratadas externamente de forma a assegurar maior independência e isenção dos resultados. De forma geral, avaliações internas, externas e mistas podem cumprir todas as funções e garantir o respeito aos princípios desde que haja intenção e uma estratégia clara de utilização dos resultados.

4.2 – Métodos e desenhos de avaliação

O tipo de método e desenho de avaliação que se pretende executar depende fundamentalmente das questões que se almeja responder com ela. A primeira distinção a ser feita diz respeito a métodos quantitativos e qualitativos. Cada um deles serve a

propósitos distintos e podem ser usados de maneira combinada para se aproveitar os pontos fortes de ambos.

4.2.1 – Métodos e técnicas qualitativas

Os métodos qualitativos são caracterizados pela valorização das percepções e interpretações feitas pelos atores a respeito dos bens e serviços públicos que constituem o objeto da avaliação. Pressupõe, segundo Deslandes (2006, p. 5) que o sucesso de uma intervenção depende dos significados atribuídos a ela pelos sujeitos sociais, que variam conforme a inserção cultural, social e econômica destes sujeitos. Portanto, a avaliação qualitativa procura compreender como se combinam expectativas e necessidades de beneficiários, gestores, políticos e outros atores com os serviços oferecidos por elas. Os métodos qualitativos podem ser utilizados para realizar qualquer tipo de avaliação (processos, resultados e impactos, ex-ante e ex-post, interna, externa e mista).

Embora haja uma diversidade de técnicas, elas têm em comum o fato de serem empregadas no contexto próprio dos indivíduos, com vistas à captação do seu comportamento real. Weiss (1998, p. 253) assinala algumas das vantagens dos métodos qualitativos: (i)"capacidade de entender o desenvolvimento dinâmico do programa de acordo com sua evolução"; (ii) "especial sensibilidade à influência do contexto"; (iii) "flexibilidade de perspectiva", entre outras.

Em geral, os estudos qualitativos são realizados num número pequeno de lugares dadas as dificuldades de analisar uma grande quantidade de dados não estruturados que precisam passar por um longo trabalho de estruturação e análise, envolvendo a identificação de padrões e temas recorrentes. O trabalho de campo de pesquisas qualitativas também costuma ser mais longo do que o de pesquisas quantitativas devido ao fato de que as técnicas empregadas de amostragem e coleta de dados demandam mais tempo para sua execução. Ao selecionar os lugares em que será realizada a pesquisa, pode-se optar por conseguir a maior diversidade possível (em termos de porte municipal -número de habitantes -, localização geográfica estadual ou regional, tradição cultural e política, indicadores socioeconômicos ou indicadores relacionados à ação que se quer avaliar - grau de cobertura no município, por exemplo) ou pode-se selecionar apenas lugares típicos e/ou não típicos.

A estratégia de pesquisa em geral utilizada nas pesquisas qualitativas são os estudos de caso, os quais podem ser usados para avaliações de resultados e de processos. No entanto, os estudos de caso não necessitam utilizar apenas dados qualitativos, sendo possível também realizar estudos de caso de caráter quantitativo.

Segundo Yin (2001, p. 24), estudos de caso podem ser utilizados com um propósito explanatório, ou seja, para responder a questões referente aos *por quês* e *como* determinadas variáveis se correlacionam. Outra característica importante dos estudos de caso, que os difere dos experimentos, é o fato de que eles não lidam com a manipulação de comportamentos sociais. Os estudos de caso em geral utilizam técnicas como a observação direta e entrevistas. Uma avaliação pode ser feita utilizando-se como estratégia metodológica a realização de múltiplos estudos de caso, com as mesmas técnicas de investigação, que gerarão resultados comparáveis. Embora não se possa generalizar estes resultados para toda a população (por exemplo, o universo de beneficiários de um programa de formação de professores em todo os municípios nos quais ele está implantado), pode-se fazer generalizações teóricas ou analíticas que permitem confirmar ou refutar hipóteses a partir de um estudo de caso.

Há uma variedade de técnicas de investigação e de referenciais teóricos que podem ser empregados com abordagens qualitativas[8]. Entrevista é uma das técnicas mais utilizadas e pode ser feita de maneira estruturada (utilizando-se a mesma série de perguntas para todos os entrevistados), não estruturada (caracterizada por ser mais livre e por intervenções menos planejadas por parte do entrevistador) ou semi-estruturada (contém perguntas fechadas e abertas). Grupo focal é uma técnica empregada quando se quer apreender o conjunto de comportamentos e percepções de um determinado grupo, geralmente composto por pessoas que possuem uma característica ou uma identidade comum (alunos, professores, homens, mulheres etc), as quais expõem seus consensos e dissensos no âmbito do grupo. Outra técnica é a observação participante, na qual o pesquisador se insere numa situação social como observador. Ele passa a fazer parte do contexto no qual vivem os sujeitos observados e coleta informações privilegiadas sobre o fenômeno que procura analisar. Por último, tem-se a pesquisa documental, que procura analisar os significados contidos em documentos diversos, quer sejam escritos ou não (imagens, músicas etc).

Por interagir de forma mais profunda com os atores sociais do que ocorre no caso das pesquisas quantitativas, as avaliações qualitativas devem ter mais cuidado com questões éticas, relativas ao anonimato das pessoas e à confidencialidade dos dados, e também com possíveis vieses advindos do envolvimento do pesquisador com os entrevistados.

---

[8] Ver Deslandes (2006, pp. 8-16).

4.2.2 – Métodos e técnicas quantitativos

Os métodos quantitativos se caracterizam pela coleta de dados que podem ser traduzidos em números e assim submetidos à análise estatística. Em geral, há uma correspondência entre métodos e desenhos, ou seja, avaliações quantitativas que utilizam desenhos experimentais e quase-experimentais rígidos ao passo que avaliações qualitativas utilizam outros tipos de desenho que não enfatizam a comparação entre grupos beneficiários e não beneficiários. Não há a intenção, em estudos qualitativos, de tratar os dados com instrumentos e testes estatísticos para estabelecer correlações entre variáveis. Em estudos quantitativos, pelo contrário, tende a haver a preocupação em gerar resultados que sejam generalizáveis para a população em estudo. Daí a necessidade de coletar dados para grandes amostras de unidades selecionadas para o programa (escolas, alunos, comunidades etc).

Visto que os dados quantitativos devem ser transformados em números e coletados em grande quantidade, é necessário que os instrumentos de coleta sejam padronizados, em geral questionários administrados pessoalmente ou a distância (via correio, telefone ou internet). Como não há muito espaço para interação entre pesquisador e pesquisado, os métodos quantitativos estão menos sujeitos à alteração que é provocada nas relações sociais e no comportamento dos sujeitos que estão sendo pesquisados devido à presença de um agente externo (o avaliador). Por outro lado, as pesquisas quantitativas geram dados objetivos, porém nelas não há margem para que os entrevistados se manifestem livremente a respeito das questões em investigação nem se captam detalhes. Assim, perdem-se informações não previstas inicialmente e que podem ser importantes para explicar o fenômeno em questão.

Os *surveys* constituem a técnica de investigação mais típica dos métodos quantitativos de avaliação. Um *survey* é um método de coleta de dados para uma amostra representativa de indivíduos por meio do qual é possível obter informações sobre fatos objetivos, bem como sobre percepções, atitudes, comportamentos, crenças e expectativas, desde que o fenômeno que se quer conhecer possa ser traduzido em variáveis e indicadores relevantes e válidos. Os *surveys* utilizam métodos padronizados de coleta de dados, nos quais todos os entrevistados são submetidos às mesmas questões. Dado que eles têm pretensão de obter resultados generalizáveis, a chance de participação de cada indivíduo da população no *survey* deve poder ser estimada, quer seja ela igual ou não.

Surveys podem ser conduzidos por vários meios: telefone, correio, internet e

entrevistas face a face. As perguntas podem ser fechadas (cujas respostas relacionam-se a categorias previamente definidas) ou abertas (nas quais o entrevistado pode se expressar livremente). Na elaboração de questionários de *survey*, é preciso ter especial atenção com a forma como a pergunta é feita (que pode afetar decisivamente as respostas), a ordem das perguntas e a quantidade de perguntas feitas, isto é, a extensão do questionário. Independentemente do tipo de respondente (beneficiários, gestores, políticos, especialistas e outros), é importante que as perguntas sejam claras e inteligíveis. Daí a importância de se realizarem pré-testes em amostras representativas da população de respondentes-alvo. Este procedimento é feito para refinar questões, torná-las mais claras e precisas, testar métodos de aplicação e conhecer as reações dos respondentes ao *survey* antes que ele seja aplicado em toda a amostra de forma que haja tempo e oportunidade para correção de erros. Outro ponto importante diz respeito ao fato de que os mesmos respondentes podem ser solicitados a responder a um *survey* em diversos momentos, por exemplo, quando se trata de um *survey* longitudinal, que pretende aferir a evolução do fenômeno em investigação (ou a sustentabilidade dos impactos de uma ação) ao longo do tempo.

4.2.3 - Desenhos

O desenho corresponde à especificação metodológica detalhada de um plano de avaliação. Após a definição das questões a serem avaliadas e do método (quantitativo e/ou qualitativo) a ser utilizado, deve-se definir então como este método será aplicado à avaliação em tela, que tipos de comparações deverão ser feitas para responder às questões selecionadas, que tipos de dados serão coletados, com que freqüência e por meio de que técnicas (*surveys*, grupos focais etc), quem serão os atores investigados (gestores estaduais e municipais, diretores de escola, professores, alunos etc) e como serão selecionados (que tipo de amostragem ou censo), quais as variáveis dependentes e independentes, além de outras especificações julgadas necessárias. Para escolher um desenho entre as diversas opções, é preciso pensar nas restrições existentes (orçamentárias, temporais, informacionais e outras) e nos interesses dos diversos atores envolvidos na avaliação.

Comparações inter-grupos. Para avaliações de resultados que objetivam isolar os efeitos de uma ação, isto é, identificar com precisão que resultados foram gerados pela ação, diferenciando-os de quaisquer outras mudanças semelhantes às almejadas pela ação que possam ocorrer na vida dos beneficiários, ocasionadas por outros fatores,

o tipo de desenho de avaliação mais robusto é o experimental aleatório. Este desenho se caracteriza por distribuir o benefício (ou serviço associado à ação) aleatoriamente entre uma determinada população homogênea em termos de características relevantes, como renda familiar per capita, local de moradia, taxa de escolaridade, influências externas etc. Desta forma, haverá um grupo de controle (formado por não beneficiários) e um grupo de tratamento (composto por beneficiários) equivalentes. Após algum tempo de implementação da ação, as diferenças identificadas poderão ser atribuídas à ação com certa segurança.

Entretanto, desenhos experimentais não são realizados com muita freqüência porque suas exigências relativas à seleção de beneficiários e não beneficiários tendem a não ser aceitas pelos gestores. Outro impedimento a este tipo de desenho é a necessidade de que a avaliação seja pensada e estruturada desde a fase anterior ao início da implementação, o que não costuma ocorrer.

Desenhos quase-experimentais podem ser realizados quando as condições para a realização de desenhos experimentais são inexistentes. Neste caso, selecionam-se grupos de comparação, formados por beneficiários e não beneficiários, com características equivalentes (sexo, cor, idade e outras características relevantes para o programa). Pode-se pensar em desenhos quase-experimentais nos quais só se têm dados para o período posterior à implementação da ação. A dificuldade reside em que, não havendo dados que assegurem a comparabilidade dos dois grupos para o período anterior à intervenção, não se pode asseverar com segurança que os resultados encontrados devem-se ao programa, e não a diferenças entre os dois grupos que a falta de dados pode haver impedido que fossem percebidas.

A utilização de dados coletados em dois momentos – antes e depois da implementação do programa –, assim como o uso de séries temporais, permite fazer inferências mais confiáveis desde que os dois grupos sejam realmente semelhantes. Caso o grupo de beneficiários tenha se inscrito por conta própria no programa, é provável que os componentes deste grupo sejam diferentes do grupo de comparação em termos de motivação, o que pode tornar a comparação menos válida e levar a uma sobre-estimação dos resultados. Em geral, usa-se uma técnica chamada emparelhamento (*matching*) para assegurar, com base em características observáveis, que os grupos são comparáveis.

Comparações intra-grupo. Existem também os desenhos que não fazem comparações inter-grupos (beneficiários e não beneficiários), porém podem implicar a

realização de comparações intertemporais (antes e depois do início da implementação da ação). No caso em que não é possível fazer comparações antes e depois, ou seja, quando a avaliação só começa depois que a ação já se iniciou, o avaliador pode fazer suposições sobre a situação inicial com base em registros administrativos, dados secundários, dados relativos a novos beneficiários ou por meio de coleta de dados junto aos próprios atores investigados a respeito de como era sua situação no período anterior à entrada no programa.

Comparações antes e depois tornam o desenho mais robusto, especialmente quando se analisa também o período durante. O problema principal reside em não saber que outros fatores podem ter afetado os beneficiários entre os dois intervalos (primeira e última coletas) que possam ter interferido na obtenção dos resultados. Quando não se levam em consideração os possíveis fatores intervenientes, pode-se estar subestimando ou superestimando os resultados. A coleta e análise de dados ao longo da implementação do programa (tanto para indicadores de processo quanto para indicadores de resultado), pautando-se na teoria do programa, pode ajudar a tornar os resultados mais confiáveis.

Também é possível formar uma série temporal, coletando dados a respeito de variáveis de interesse para períodos anteriores e posteriores ao início da ação. Dados relativos a períodos anteriores ao início de um programa específico podem ser os dados coletados sistematicamente a respeito de variáveis educacionais por instituições como o INEP, que certamente constituem um suporte importante para avaliações. Quanto à coleta de dados para períodos muito posteriores ao término de uma ação, embora ela raramente ocorra, sua importância está em poder avaliar a sustentabilidade de uma intervenção, assim como seus efeitos a longo prazo.

Outra possibilidade é fazer comparações internas, utilizando-se a heterogeneidade no tratamento contida no programa como base para comparações. Este procedimento é conhecido como desenhos de “dose-resposta” (Weiss, 1998, p. 195) e é baseado em métodos utilizados na medicina. O termo “heterogeneidade dos tratamentos” refere-se às diferentes instituições ou aos diversos níveis de governo envolvidos na prestação de bens ou serviços entregues pelo programa, aos diferentes contextos em que atuam tais instituições e às diversas metodologias com as quais trabalham os prestadores de serviço. Outro elemento importante a se considerar é a heterogeneidade das comunidades às quais se dirige o programa. Fatores relativos às comunidades que importam para a avaliação são: exposição prévia à intervenção;

diferentes perfis de beneficiários, em termos de idade, expectativas, experiência prévia com o objeto do programa (por exemplo, cursos de alfabetização) e lugar onde será realizada a intervenção (zona rural/urbana, metrópoles, pequenos municípios etc).

Programas que contêm estes tipos de heterogeneidade podem ser avaliados por meio de comparações que levem em conta estas diferenças para fazer afirmações acerca dos resultados. A idéia básica é que pessoas que têm acesso a uma "dose" maior do programa apresentarão melhores resultados. Neste caso, é preciso checar se realmente a comparação está sendo feita entre beneficiários com as mesmas características (ou seja, não se pode comparar pessoas que recebem uma dose maior porque valorizam mais o programa e por isso o procuram mais de forma espontânea com pessoas desmotivadas que recebem uma dose menor porque não se interessam pelo programa). A mesma idéia pode ser aplicada à comparação entre beneficiários que têm acesso a diferentes serviços no âmbito de um mesmo programa (por exemplo, pessoas que são alfabetizadas por meio de diferentes metodologias).

4.2.3 – Tipos de avaliações e métodos utilizados

Avaliações de processo usam métodos quantitativos ou qualitativos, podendo-se combinar ambos os tipos para aproveitar as vantagens comparativas de cada uma. Métodos quantitativos são mais facilmente empregados quando os processos são claramente identificados e institucionalizados. Quando não se dispõe de muita informação sobre tais processos, pode-se considerar fazer uma avaliação qualitativa. A partir dos resultados desta avaliação qualitativa, pode-se pensar em indicadores quantitativos ou qualitativos de capacidade institucional e de eficiência e eficácia da gestão para monitoramento e para outras avaliações. Para medir a eficiência da gestão, realizam-se também estudos de custos, que objetivam determinar os custos dos bens e serviços providos e analisar a relação custo-qualidade ou custo-efetividade de diferentes modelos de intervenção. Desta forma, combinam-se avaliações de processos com avaliações de resultados, que se destinam a aferir a eficácia e a efetividade das ações.

Avaliações de impacto se prestam a aferir os resultados finais atingidos pelo programa sobre determinadas dimensões. Em geral, elas são conduzidas por meio de métodos quantitativos, com desenhos quase-experimentais, que implicam a comparação entre grupos de tratamento e controle, análise dos dados utilizando recursos estatísticos e econométricos. A necessidade de um contrafactual, ou seja, de saber como pessoas com perfil semelhante ao dos beneficiários se comportam em relação a determinados aspectos que deveriam ser afetados pelo programa constitui um imperativo das

avaliações de impacto. No entanto, nem todos os programas precisam passar por este tipo de avaliação visto que, devido às suas especificidades, elas tendem a acarretar altos custos. Avaliações de impacto rigorosas devem ser realizadas com desenhos experimentais ou quase-experimentais e requerem representatividade, o que implica a composição de grandes amostras.

Também é possível avaliar resultados por meio de avaliações menos dispendiosas e com outros tipos de métodos e desenhos, diferentes daqueles utilizados por uma avaliação de impacto tradicional. Segundo Weiss (1998, p. 86), avaliações de resultado quantitativas privilegiam a mensuração de resultados ao passo que avaliações qualitativas podem revelar, por exemplo, o significado dos resultados para os beneficiários. A triangulação de métodos quantitativos e qualitativos é aconselhável visto que permite maximizar os pontos fortes de cada um deles e cruzar resultados de modo a obter-se uma visão mais completa do fenômeno em investigação, debruçando-se, assim, simultaneamente, sobre insumos, processos e resultados. Outra alternativa é realizar uma série de estudos seqüenciais de escopo menor, em lugar de um grande estudo, que utilizem diferentes métodos e que possam responder às questões desejadas paulatinamente, priorizando-se aquelas consideradas mais relevantes e contemplando aquelas que surgem diante dos desafios postos pela implementação da ação.

5. Disseminação dos resultados

O DAIE tem feito ampla divulgação dos resultados de suas avaliações. Há uma linha editorial do MEC chamada Coleção Educação para Todos pela qual já foram publicados diversos volumes relativos à área de Avaliação e Monitoramento[9]. Além das publicações, outros meios adotados pelo DAIE para disseminar seu trabalho são seminários e oficinas de discussão sobre Avaliação e Monitoramento, para os quais são convidados pesquisadores em geral, especialistas em avaliação, gestores dos três níveis de governo, representantes de ONGs e demais organizações da sociedade civil envolvidas na gestão das políticas da SECAD. Entende-se a ampla divulgação dos seus produtos como uma das boas práticas desta organização, não apenas devido ao respeito ao compromisso com *accountability* que toda a administração pública deve ter, mas também por causa dos efeitos positivos ensejados por ela.

A publicação é uma etapa imprescindível da produção científica dado que é a

[9] Ministério da Educação: Coleção Educação para Todos, volumes 18-25.

partir dela que se socializam os novos conhecimentos e se realiza o intercâmbio entre pesquisadores que possibilita o crescimento do campo. A publicação das pesquisas avaliativas é, portanto, o meio pelo qual os pesquisadores obtêm reconhecimento de seus pares, ao prover-lhes as informações necessárias para a replicação de seus trabalhos, alimentando assim o debate sobre seus achados. Desta forma, a avaliação ganha notoriedade, o que potencializa o seu uso na gestão. Na medida em que a comunidade científica e a sociedade em geral se apropriam de uma pesquisa avaliativa, gestores e políticos não poderão facilmente ignorá-la.

Outra iniciativa do DAIE neste sentido é a criação de uma Rede de Avaliadores dedicados às políticas públicas implementadas pela SECAD. Esta Rede, que se encontra em processo de formação, é um instrumento importante para fomentar o desenvolvimento do campo de avaliação de programas educacionais, especialmente dos programas destinados a promover a diversidade na educação, dado seu caráter inovador. Por meio dela, o DAIE tem ampliado o número de seus interlocutores e os incentivado a debruçar-se sobre os temas e problemas que são objeto de intervenção desta Secretaria. Desta forma, pesquisadores e especialistas em avaliação adquirem a oportunidade de interagir com tomadores de decisão no âmbito do Estado, compartilhar conhecimentos, realizar pesquisas avaliativas e assim contribuir para o aumento da eficácia das políticas públicas educacionais.

6. Comentários finais

A definição de conceitos e metodologias de avaliação e dos métodos e procedimentos utilizados pelo DAIE neste texto constitui mais uma iniciativa no sentido da progressiva institucionalização da avaliação no âmbito da SECAD. Além disso, este texto integra uma publicação que faz um esforço de sistematização da experiência do DAIE até o momento e de apresentação de como ele será continuado, ou seja, como o departamento pretende consolidar sua agenda nesta gestão. Acredita-se que o trabalho em conjunto com os demais departamentos da SECAD será facilitado por esta publicação, que traz uma clara e completa exposição dos interesses e atribuições do DAIE. Desta forma, contribui-se para a desmistificação da visão de avaliação como obstáculo ou como mecanismo reducionista dos resultados obtidos pelas ações a simples números ou indicadores pouco compreensivos. Espera-se que, a partir deste texto e da experiência prática de contato com avaliação que se seguirá, solidifique-se o processo de sedimentação de uma cultura de avaliação integrada à gestão.

Referências Bibliográficas:

DESLANDES, Suely. *Metodologia de avaliação III: introdução aos métodos qualitativos.* Brasília: ENAP, 2006.

DRAIBE, Sônia. "Avaliação de implementação: esboço de uma metodologia de trabalho em políticas públicas", em: BARREIRA, Maria e CARVALHO, Maria (Orgs). Tendências e perspectivas na avaliação de políticas e programas sociais. São Paulo: IEE/PUC-SP, 2001.

Manual de Avaliação do Plano Plurianual 2004-2007 Exercício 2007, Ano Base 2006. Ministério do Planejamento, Secretaria de Planejamento e Investimentos Estratégicos. Brasília, 2006.

MOREIRA, Tadiana Maria Alves. *Estudo de caso da avaliação da descentralização das ações programáticas de hanseníase*. Tese de Doutorado. Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública; 2002.

PÓ, Marcos Vinícius e ABRÚCIO, Fernando. "Desenho e funcionamento dos mecanismos de controle e accountability das agências reguladoras brasileiras: semelhanças e diferenças", em: *Revista de Administração Pública.* Rio de Janeiro:40 (4): pp. 679-98.

PAES DE BARROS, Ricardo e CARVALHO, Mirela. *Avaliação: o que é e para que serve?* Brasília: ENAP, 2006.

RUA, Maria das Graças. A avaliação no ciclo da gestão pública. http://www.pr.gov.br/sepl/avalciclogestpubl.doc

SANTOS JÚNIOR, Orlando Alves. *Democracia e governo local: dilemas da reforma municipal no Brasil*. Rio de Janeiro: Revan, 2001.

WEISS, Carol. *Evaluation: methods for studying programs and policies.* Harvard University: 1998.

WHOLEY, Joseph, HATRY, Harry e NEWCOMER, Kathryn (Eds.) *Handbook of practical program evaluation*. São Francisco: Jossey-Bass, 2004.

YIN, Robert. *Estudo de caso: planejamento e métodos.* Porto Alegre: Bookman, 2001.

Box – Modelo Lógico

O Modelo Lógico é um instrumento teórico que serve como subsídio à elaboração do Plano de Avaliação. No Canadá, segundo Moreira (2002), todas as avaliações devem ser precedidas de modelos teóricos semelhantes ao Modelo Lógico, cujo propósito é explicitar a teoria do programa associada à intervenção que se pretende avaliar. É uma metodologia que parte da contextualização do problema, identificando as causas, conseqüências e meios adotados pela ação como proposta de resolução do problema apontado. Seu modelo de apresentação segue o fluxo a partir do problema, suas causas e conseqüências, detalhando também os componentes da ação (insumos, atividades e resultados de curto, médio e longo prazo). Trata-se de um quadro referencial que informa os avaliadores acerca do funcionamento básico da ação de forma a proporcionar-lhes as informações necessárias para desenhar a avaliação, tais como: que resultados o programa pretende atingir, que atividades têm sido implementadas para alcançar os resultados esperados, em que medida a obtenção dos resultados afetará o problema diagnosticado etc.

O Modelo Lógico pode ser revisado a qualquer tempo à medida que novas informações são coletadas ou que ocorrem modificações no desenho do programa. Algumas das vantagens do uso deste instrumento são:

- Definição do grau de avaliabilidade da ação;
- Auxílio na definição de conceitos relacionados ao programa, no planejamento e na comunicação externa e interna sobre o programa;
- Padronização do entendimento sobre o programa entre equipe técnica e outros participantes;
- Explicitação de possíveis incoerências lógicas entre relações causais subjacentes ao desenho do programa.

Em geral, o fluxo esquemático do Modelo Lógico contém os seguintes elementos: Causa – Problema – Programa – Público-alvo – Objetivo geral – Objetivo específico – Recursos – Ações – Atividades – Produtos – Resultados Intermediários – Resultados finais. Os dados utilizados na elaboração do quadro podem advir de diversas fontes, tais como: literatura científica, entrevistas com gestores, especialistas e outros atores envolvidos com o programa, documentos oficiais sobre o programa etc.