

## 2

# Avaliação de Programas

definição, tipologias e etapas  
para sua elaboração

*A avaliação de programas produz evidências sobre demandas sociais, características de públicos-alvo, implementação de atividades e seus resultados e impactos. Para tanto, requer clareza de objetivos a investigar e escolha adequada do método a empregar. Pior do que não ter uma avaliação do programa é dispor de um estudo apressado, mal-especificado ou mal-conduzido, pior ainda se elaborado por uma instituição respeitada. Não bastasse não trazer conhecimento relevante, pode informar incorretamente a decisão sobre mérito e aspectos cruciais do programa.*

**A**mpliação do escopo e escala das políticas e programas sociais no Brasil, em especial a partir da Constituição Federal de 1988, tem reclamado o aprimoramento da capacidade de formulação e avaliação de programas públicos no país. De modo geral, os estudos de dimensionamento de demandas sociais, os diagnósticos propositivos para intervenções específicas, os sistemas de indicadores de monitoramento e as pesquisas de avaliação de implementação e de resultados de programas vêm se tornando mais consistentes e presentes no cotidiano dos gestores públicos.

O Brasil segue o caminho trilhado por outros países em décadas passadas, a passos mais largos em algumas áreas setoriais de políticas públicas – Saúde, Educação, Desenvolvimento Social – e esferas administrativas – governo federal, alguns estados pelo país e municípios de maior porte. Como nos países mais desenvolvidos, começa-se a constatar indícios de que a cultura de avaliação de programas está se institucionalizando, depois de mais de duas décadas de construção, com produção de estudos e experiências consolidadas, marcos normativos em definição, comunidades de profissionais especializados se formando e com oferta regular de formação na área de M&A em universidades e centros de pesquisa aplicada.

Nos Estados Unidos e em alguns países europeus, os primeiros esforços sistemáticos em avaliação de programas sociais começaram a se estruturar a partir da década de 1930, como desdobramento do papel crescente do Estado no financiamento de serviços sociais e respondendo aos efeitos de pauperização social e desemprego deflagrados pela crise de 1929 (Rossi *et al.*, 2004). Mas é na reconstrução do pós-guerra e com a criação de políticas sociais mais abrangentes que a cultura de monitoramento e avaliação de programas iniciou sua concretização. A avaliação de programas, que se concentrava, inicialmente, nas áreas de educação (sobretudo na investigação de programas de alfabetização), de saúde pública (nas ações de combate à mortalidade por doenças infectoparasitárias) e de qualificação profissional, passa a abranger áreas como prevenção da violência juvenil, planejamento familiar, nutrição, programas habitacionais, desenvolvimento rural.

A prática da avaliação de políticas e programas públicos – assim como os instrumentos de planejamento e programação orçamentária – viria a receber ainda mais apoio no Governo Lyndon Johnson, quando da estruturação de vários programas

sociais em nível federal para combate à pobreza e desigualdade racial – o plano *Great Society*. Era necessário garantir melhor gestão dos programas públicos perante os gastos crescentes em políticas sociais, processo que se manteve nos governos seguintes, com o reforço técnico proporcionado pela estruturação de cursos de pós-graduação na área, com forte ênfase em técnicas de pesquisa social (quantitativas e qualitativas), o que abriu a possibilidade para a criação de uma comunidade profissional de avaliadores.

A partir da década de 1980, ao contrário do que se poderia imaginar, com a volta de governos republicanos conservadores e a mudança da agenda pública – redução dos gastos em programas sociais, desregulamentação e descentralização da prestação de serviços públicos – a prática da avaliação não esmoreceu, espalhando-se para governos estaduais, municipais e organizações não governamentais. Afinal, mesmo um “estado regulador” requer instrumentos de medição de resultados para acompanhar a prestação dos serviços financiados com recursos públicos. Também precisa dispor de informação, com alguma base técnica e transparência, para justificar a descontinuidade de programas sociais “ineficazes” e redirecionar os gastos para outras áreas (como para defesa militar, no governo Reagan).

Essa recensão histórica da avaliação nos EUA mostra que, ao longo de algumas décadas, a avaliação de programas deixou de ser preocupação apenas de cientistas sociais, em temas de pesquisas acadêmicas nas universidades e ganhou projeção e interesse na comunidade de gestores públicos como instrumento para aprimoramento dos serviços nesse âmbito. Processo semelhante verificou-se em países europeus nesse período, o qual culminou com a formação de larga infraestrutura de formação profissional e apoio à realização de estudos de avaliação<sup>2</sup>.

Com alguma defasagem, o Brasil está se aproximando desse estágio de institucionalização da avaliação, como demonstram a oferta de cursos de capacitação e pós-graduação nas universidades, a criação de associação de profissionais de avaliação e, em particular, a produção de estudos nos centros de pesquisas e no próprio setor público (Joppert *et al.*, 2011; Jannuzzi, 2014). De fato, há um número significativo de pesquisadores – com atuação em programas de mestrado e também de doutorado – com área de concentração ou linhas de pesquisas em Avaliação de Políticas e Programas, como mostra o Diretório de Grupos de Pesquisa do CNPq<sup>3</sup>. Há uma produção regular de trabalhos em M&A, apresentados em congressos e publicados

2. Vide, por exemplo, as publicações, eventos e sites das associações profissionais ou acadêmicas de avaliação nesses países, entre os quais: American Evaluation Association ([www.eval.org](http://www.eval.org)); Canadian Evaluation Society ([www.evaluationcanada.ca](http://www.evaluationcanada.ca)); European Evaluation Society ([www.europeanevaluation.org](http://www.europeanevaluation.org)) e Société Française d’Evaluation ([www.afe.asso.fr](http://www.afe.asso.fr)).

3. Em meados de 2015, havia mais de trezentos grupos registrados no diretório com temas de avaliação de programas. Vide <<http://lattes.cnpq.br/web/dgp>>.

nas várias revistas científicas existentes nas Ciências Sociais Aplicadas, Educação, Saúde Pública e Economia no país<sup>4</sup>. Além da *Revista de Administração Pública*, *Planejamento e Políticas Públicas* e *Revista do Serviço Público*, que têm trazido volume crescente de artigos na temática, há, pelo menos, sete revistas técnicas e/ou científicas criadas para dar maior visibilidade a tais temas: a *Revista Brasileira de Monitoramento e Avaliação*, a *Revista de Avaliação de Políticas Públicas*, a *Estudos em Avaliação Educacional*, a *Ensaio – Avaliação e Políticas Públicas em Educação*, a *Revista de Avaliação em Educação Superior*, *Boletim de Políticas Sociais* e *Cadernos de Estudos Desenvolvimento Social em Debate*.

A diversidade temática do conjunto de pesquisas regulares do IBGE, os estudos e livros do Ipea, as publicações do Ministério da Saúde, os levantamentos de dados educacionais do Instituto Nacional de Estudos e Pesquisas Educacionais (Inep), as pesquisas disponibilizadas pela Secretaria de Avaliação e Gestão da Informação do Ministério de Desenvolvimento Social e Combate à Fome (SAGI/MDS) são outras demonstrações claras de como as instituições brasileiras vêm respondendo aos desafios de dispor de informação qualificada para aprimoramento de suas políticas e programas públicos.

Tais avanços são acompanhados também de limitações ou disfunções. A prematuridade na encomenda de estudos avaliativos de resultados e impactos, quando se sabe que o programa ainda se encontra em fase de implantação ou com problemas de gestão, assim como a antecipação de avaliações externas, com natureza mais de auditoria de desempenho, em detrimento de avaliações de caráter mais formativo e voltadas ao aprimoramento de programas, são problemas mais frequentes do que se poderia esperar. Não é raro também se encontrarem situações em que são aportados volumes significativos de recursos em pesquisas de avaliação, com desenhos complexos e amostras extensas, em detrimento de desenvolvimento de painéis de indicadores de monitoramento que – com todas as limitações de cobertura, mas com a tempestividade devida – permitiriam identificar boa parte dos problemas em tempo de serem corrigidos. Vários estudos avaliativos focados no processo de implementação dos programas acabam tendo baixo poder de generalização por causa das lacunas de conhecimento sobre as atividades desenvolvidas e a respeito do papel dos agentes institucionais envolvidos nos programas, seja pelas limitações ou inexistência de sistemas de informações de monitoramento, seja pelas decisões metodológicas quanto às técnicas, amostras e casos investigados, muitas vezes escolhidos mais em função dos prazos e recursos disponíveis que pelas perguntas as quais se quer responder.

4. Ver, sobre esse assunto, a programação dos seminários regulares das principais associações científicas como a Anpocs, Abep, SBS, Abrasco, Anpec, Anpad, entre outros.

Há também situações concretas de avaliações mal concebidas, apressadas ou mal-executadas de programas públicos, as quais produzem informações que, infelizmente, assentadas na reputação anterior da instituição avaliadora, acabam virando “verdades” incontestes – e factoides na imprensa – o que leva a questionamentos quanto ao mérito do programa, abala o empenho das equipes envolvidas e conduz a decisões equivocadas na implementação dos programas (Worthern *et al.*, 2004). São avaliações que, ao contrário do que esperariam seus proponentes e executores, embebidos por certo tecnocratismo ingênuo, não contribuem com soluções, mas introduzem novos problemas no desenho e operação dos programas. Enfim, há vários exemplos de pesquisas de avaliação que, mesmo motivadas por preocupação legítima e meritória com a eficácia e efetividade do gasto público, revelam, infelizmente, resultados triviais, metodologicamente questionáveis ou com baixa aplicabilidade para aprimoramento dos programas. Nesse quadro, em que pese o esforço de organização de base de dados e produção de estudos sobre políticas e programas no país, não há como não reconhecer a procedência de certo desencanto de técnicos e gestores com os estudos avaliativos, pelos resultados inconclusivos, morosos e, em alguns casos, irrelevantes a que chegam, problemas apontados já há um bom tempo por Cotta (1999).

Este livro e, em particular, este capítulo, é uma contribuição para avançar positivamente nessa agenda de M&A no país, ao prover uma sistematização de conceitos e técnicas com base na experiência concreta vivenciada no Brasil. É o que se começa a apresentar, nas seções seguintes.

## Avaliação de programas: uma definição operacional

A avaliação de programas tem recebido as mais diferentes definições na literatura especializada, segundo os diversos modelos conceituais, paradigmas teóricos e linhas de pesquisa da Ciência Política, Ciências Sociais, Economia e Administração Pública, tomando-se a avaliação em perspectiva mais geral, como componente integrante da análise de políticas públicas, como instrumento de sistemas de monitoramento e avaliação de programas governamentais ou, em uma concepção mais restrita, avaliação como um tipo particular de investigação empírica acerca de programas e projetos sociais, como as avaliações de impacto experimental ou quase-experimental<sup>5</sup>.

A pluralidade de definições remete, por sua vez, a diferentes formas de classificação das pesquisas e estudos de avaliação e produz uma taxonomia de pouca

5. Vide McDavid e Hawthorne (2006); Mackay (2007); Worthern *et al.* (2004); Rossi *et al.* (2004).

utilidade para quem quer entender os resultados de avaliação e, possivelmente, mesmo para aqueles que as produzem. Combine-se essa compulsão classificatória – motivada por certo academicismo ou disputas de espaço no campo – com as várias técnicas de pesquisa social empírica disponíveis e o caos está formado: avaliação orientada a objetivos, avaliação multissítios, avaliação focada na decisão, avaliação participativa, avaliação orientada a usuários, avaliação responsiva a agentes etc. Novamente, então, é preciso reconceituar-se o termo avaliação, já que cada uma dessas modalidades acaba remetendo a uma finalidade diferente.

Evitando enveredar-se pela prolixidade terminológica da área, apresentando as variadas definições de avaliação, as suas exaustivas formas de categorização e as inúmeras estratégias metodológicas existentes para conduzi-las, esta sessão e as que se seguem procuram trazer os conceitos, categorias essenciais e processos de trabalho envolvidos nesse campo. Com os termos básicos assimilados, pode ficar mais fácil entender as nuances e os novos tipos de estudos propostos na área.

Assim, em uma definição pragmática e aplicada, avaliação refere-se ao esforço analítico de produzir informação e conhecimento para desenho, implementação e validação de programas e projetos sociais, por meio de abordagens metodológicas interdisciplinares da pesquisa social, com a finalidade de aprimorar a gestão das intervenções, seja no cumprimento dos seus objetivos (eficácia), nos seus impactos mais duradouros e abrangentes em termos de públicos e dimensões sociais alcançados (efetividade), seja nos custos condizentes com a escala e complexidade da intervenção (eficiência). Dessa maneira, configuram-se como avaliação as pesquisas para dimensionamento e entendimento dos determinantes de problemas sociais, para dimensionamento e caracterização de públicos-alvo de possíveis programas, para investigar as dificuldades de desenvolvimento de determinadas atividades previstas na implementação de um programa, assim como os resultados, efeitos mais abrangentes e custos destes (Figura 2.1).

Em tal definição, avaliação não é entendida tão somente como uma investigação com métodos validados cientificamente para analisar diferentes aspectos sobre um programa – o que se constituiria em uma investigação de cunho acadêmico –, mas um levantamento consistente, sistemático e replicável de dados e produção de conhecimento para aprimoramento da intervenção programática, que versa sobre: as características essenciais do contexto de atuação e os públicos-alvo; o desenho e os arranjos de implementação; os custos de operação; os resultados de curto prazo; os impactos sociais e de mais longo prazo de um programa. Enfim, na definição aqui advogada, avaliação tem a finalidade de produzir evidências, compilar dados e sistematizar estudos que contribuam para o aperfeiçoamento de programas e projetos sociais e à consecução de seus objetivos (Vaistman; Paes-Sousa, 2011).

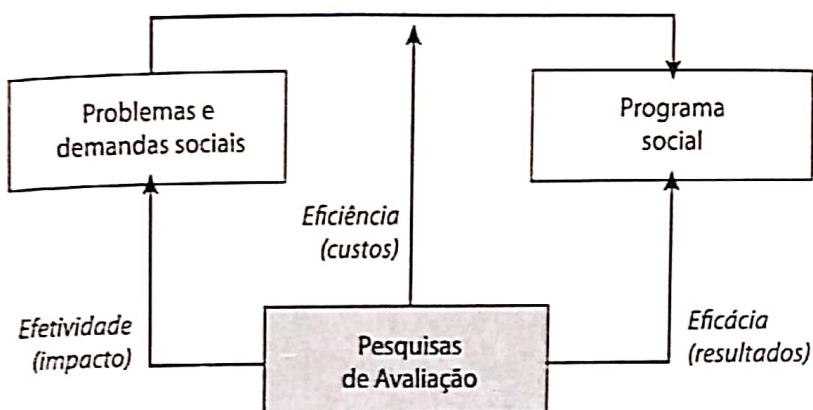


Figura 2.1. Pesquisas de avaliação de Programas Sociais.

A definição de avaliação de programas também remete à necessidade de abordagem interdisciplinar na produção de informação e conhecimento sobre os problemas investigados. A avaliação de programas não é campo de pesquisa exclusivo de uma disciplina científica em particular. Os conceitos, categorias analíticas, paradigmas e modelos interpretativos, técnicas das Ciências Sociais, Economia, Antropologia, Demografia, Estatística, Ciência Política, Administração Pública, contribuem para desvelar as problemáticas sociais e a busca de soluções para os desafios dos programas públicos. Afinal, pobreza, desigualdade, desempenho escolar e mortalidade infantil são questões sociais multideterminadas, isto é, determinadas e influenciadas por uma série de dimensões sociais mais amplas e estruturais. Programas são empreendimentos complexos, que envolvem a contratação de pessoal técnico, disponibilidade de instrumentos, adequação de equipamentos públicos, alocação de recursos monetários, promoção de capacitação, de forma coordenada, no tempo e no território. Mapear, pois, as dimensões sociais mais relevantes à formulação de um programa, seus problemas operacionais, restrições e impactos requer uma equipe de avaliadores com diferentes formações acadêmicas e profissionais. Além de especialistas acadêmicos nas temáticas investigadas, as equipes de avaliação precisam dispor também de profissionais que já tiveram experiência na gestão e operação de programas correlatos.

A definição apresentada remete, ainda, aos objetivos finalísticos das avaliações. Para que, afinal, se faz avaliação de programas sociais? Da perspectiva que aqui se defende, as avaliações constituem instrumentos para aprendizagem organizacional com a finalidade de aprimorar a gestão e o atingimento dos resultados de programas sociais. Há outras perspectivas finalísticas da Avaliação, discutidas no Capítulo 5, como o julgamento de mérito dos programas para fins de dotação de recursos ou responsabilização de suas atividades. Na sua concepção como mecanismo de aprendizagem organizacional, a avaliação produz insumos relevantes para análise da eficiência no uso

dos recursos, eficácia em atingir os objetivos e efetividade na mudança social por ele proporcionada (Draibe, 1999). Trazem-se, assim, subsídios para o julgamento técnico-político da relevância e sustentabilidade do programa no tempo.

Eficiência é um atributo relacionado ao custo que se incorre na produção dos resultados do programa. Em geral, a eficiência de um programa é avaliada em termos comparativos com outras soluções desenhadas para mesma problemática. A eficácia está relacionada ao cumprimento dos objetivos explicitados do programa. A efetividade ou impacto diz respeito aos efeitos sociais mais abrangentes do programa, antecipáveis ou não, para além de seus objetivos, medidos no seu público-alvo ou em outros segmentos da sociedade e verificáveis num prazo razoável, no qual se possa garantir algum nível de atribuição ou associação com o programa.

A avaliação da eficácia de um programa social, certamente, pressupõe uma análise da cobertura, grau de focalização do público-alvo alcançado e qualidade das entregas, além de demandar que se verifique se outros objetivos definidos no programa foram atingidos. A análise da eficácia de um programa de combate à fome, por exemplo, pressupõe a verificação da cobertura da população mais vulnerável e a queda da desnutrição infantil e insegurança alimentar nessa população. O programa pode estar reduzindo a desnutrição infantil em segmentos populacionais específicos, mas não na população em geral. Sua eficácia está restrita aos públicos-alvo focados, mas com baixa cobertura populacional. A qualidade dos alimentos distribuídos ou dos serviços prestados para que os beneficiários possam desfrutar de maior segurança alimentar é outro aspecto a considerar.

Da mesma forma, a avaliação da efetividade envolve a investigação de impactos sociais mais abrangentes proporcionados pelo programa. Um programa de transferência de renda pode ser eficaz na redução da pobreza monetária, mas sem maiores efeitos para outras dimensões sociais. Já um programa dessa natureza, que envolva também ações de provimento de vagas em escolas e atendimento à saúde a crianças e mães, como o Programa Bolsa Família, pode gerar efeitos conjugados no aumento do atendimento escolar e cobertura vacinal. Se se trata de um programa com desenho que contemple recursos de acordo com o número de filhos, seus efeitos na desnutrição infantil também devem ser mais intensos.

Em um programa de qualificação profissional, pode-se avaliar a eficácia de sua capacidade de mobilizar alunos para programas de intermediação de emprego e de acesso a microcrédito, bem como a sua efetividade em viabilizar mobilidade ocupacional ou aumento de salários de seus alunos egressos. Na dimensão da efetividade, pode estar também contemplada a equidade na cobertura de públicos com menor escolaridade ou com maiores dificuldades de acessar vagas em tais programas. Afinal, programas públicos de qualificação profissional não deveriam proporcionar



oportunidades também para pessoas desempregadas, mais pobres, residentes em municípios mais distantes? Ou deveriam continuar concentrados nos municípios e regiões com mercados de trabalho mais dinâmicos, voltados apenas a segmentos de trabalhadores já inseridos no mercado formal de trabalho?

A eficiência de uma ou outra estratégia para mitigar a fome de crianças pode ser medida por meio de indicadores de custo da produção dos resultados. Denominado de custo-efetividade, esse indicador de eficiência permite apurar qual programa reduziu mais a desnutrição infantil por valor orçamentário aportado. Naturalmente, para cálculos consistentes de custo-efetividade, é necessária uma contabilidade de custos muito detalhada e precisa – difícil pela natureza complexa e integrada de programas sociais –, além da identificação de indicadores de resultados comparáveis pelas diferentes estratégias. Para tornar ainda mais complexa tal apuração, é preciso considerar que existem custos não facilmente tangíveis na operação de programas, como os “custos políticos” de construção dos diferentes arranjos institucionais em que se baseiam. Além disso, mesmo que programas sociais diferentes produzam resultados próximos para cada real aportado neles, cada programa gera impactos sociais e externalidades negativas que deveriam fazer parte da avaliação. Ademais, há programas em que a eficiência é uma consideração secundária diante da normatividade do direito social assegurado ou da justificativa social de sua existência como programas de distribuição gratuita de remédios contra Aids, transporte público ou transferência de renda para pessoas com deficiência, distribuição de cestas de alimentos para população em áreas remotas, para citar alguns exemplos. Na análise de eficiência, é preciso considerar que existem níveis básicos de prestação de serviços, previstos na Constituição Brasileira ou em normativa particular. Postos de saúde, escolas, centros de referências da assistência social, delegacia de polícia, corpo de bombeiros e unidades de forças armadas precisam existir, para provimento de serviços básicos de cidadania e segurança. Ainda assim, para assegurar a eficiência, é necessário garantir a economicidade – busca de soluções menos dispendiosas – como princípio na alocação de recursos públicos e níveis crescentes de produtividade dos serviços.

Atestadas a eficiência, eficácia e efetividade de um programa, há fortes indicações de sua relevância ou mérito, assim como de sua sustentabilidade. O mérito ou relevância do programa estão relacionados à sua capacidade de oferecer uma solução consistente para o problema social para o qual ele foi desenhado. A sustentabilidade do programa refere-se às condições orçamentárias e operacionais de sua continuidade. Assim, avaliações que demonstrem que os objetivos do programa estão sendo cumpridos, a custos justificáveis, produzindo efeitos sociais mais abrangentes, tendem a se legitimar na sociedade e dentro do governo, contribuindo para garantir os recursos e arranjos operativos para sua continuidade.

## Tipologia de avaliação de programas: o essencial a saber

Os livros clássicos e manuais de avaliação são pródigos em apresentar diversas formas de classificação das pesquisas e estudos de avaliação, muitas vezes com pouca utilidade instrumental para a análise de programas públicos. Worthern *et al.* (2004) propõem, por exemplo, uma classificação que diferencia as pesquisas de avaliação segundo seus protagonistas. Imas e Rist (2009), por seu turno, apresentam a classificação de avaliação segundo combinação de propósitos, públicos investigados e técnicas de pesquisas empregadas.

Para avaliações de programas públicos, há algumas poucas classificações essenciais. A mais básica e simples é a categorização segundo o momento de sua realização em relação ao programa: *avaliação ex ante*, realizada antes do início de um programa, para caracterização do público-alvo e dimensionamento da problemática social que será o foco da intervenção; e *avaliação ex post*, para análise comparativa das mudanças ocorridas no público-alvo e aferição de resultados potenciais do programa. Adicionalmente, nessa tipologia, haveria a *avaliação de processos*, voltada ao exame de rotinas operacionais do programa, para identificação de gargalos e problemas de implementação (Cohen; Franco, 1994).

Uma tipologia clássica apresentada por Imas e Rist (2009) é a que classifica as avaliações em *prospectivas*, destinadas a analisar a factibilidade de sucesso do programa, com base na proposta de seu desenho lógico de atividades e agentes envolvidos; em avaliações *formativas* (ou de processo), voltadas a investigar os problemas de implementação do programa, com a finalidade de produzir insumos para solucioná-los e em avaliações *somativas* (ou de resultados), dirigidas à apreciação dos resultados e impactos dos programas, com vistas a seu julgamento de mérito, continuidade ou descontinuidade (Figura 2.2).

Outra categorização clássica de avaliações é segundo o agente avaliador do programa: *avaliação externa, avaliação interna, mista e participativa*. Em tese, avaliações com equipes externas podem garantir maior independência técnica quanto à condução do estudo avaliativo do programa que as realizadas internamente. Entretanto, nem sempre *equipes externas de avaliação* reúnem os atributos idealizados de independência, competência técnica, idoneidade e são comprometidas com uma visão multidisciplinar de avaliação de programas, que assegure a desejada credibilidade – e apropriação – dos resultados às equipes de gestores e técnicos de programas. *Equipes internas* detêm, em geral, conhecimento sobre os problemas e as dificuldades mais prementes que afetam os programas públicos, mas precisam de apoio técnico de consultores especializados e pesquisas de campo para entender suas

causas e estratégias de resolução. Com equipes *mistas*, garante-se a sinergia entre conhecimento interno de gestão e especialidade técnica externa, o que gera produtos com maior potencialidade de apropriação, não apenas pela adequação às demandas mais relevantes para avaliação como também pela legitimidade conferida pelo envolvimento inicial dos agentes internos ao programa.

Em algumas situações, menos frequentes na avaliação de programas públicos, desenvolvem-se abordagens *participativas* – ou ‘quase-participativas’ –, quarto tipo de avaliação segundo agente avaliador, com participação direta ou com algum nível de interação participativa do público-alvo e dos agentes operadores no processo de coleta e sistematização de informação. Esse tipo de avaliação é empregado quando se procura engajar os beneficiários, usuários e técnicos na concretização dos objetivos do programa. É como uma pesquisa-ação, em que a obtenção de informações é tão importante quanto promover a busca de soluções para as dificuldades do programa, como em projetos de economia solidária, de cooperativismo e autoconstrução de moradias. Em algumas situações, é a única estratégia de coleta, perante as dificuldades de abordagem de problemáticas sociais complexas ou de públicos refratários às estratégias mais convencionais, como as realizadas com população em situação de rua, com jovens em situação de drogadição, com comunidades quilombolas isoladas, entre outras. Naturalmente, nesse tipo de avaliação, os parâmetros clássicos de distanciamento do objeto, imparcialidade dos sujeitos e replicabilidade dos resultados, que orientam as metodologias convencionais cedem lugar ao potencial efeito engajador da abordagem participativa na busca de soluções para os problemas identificados.

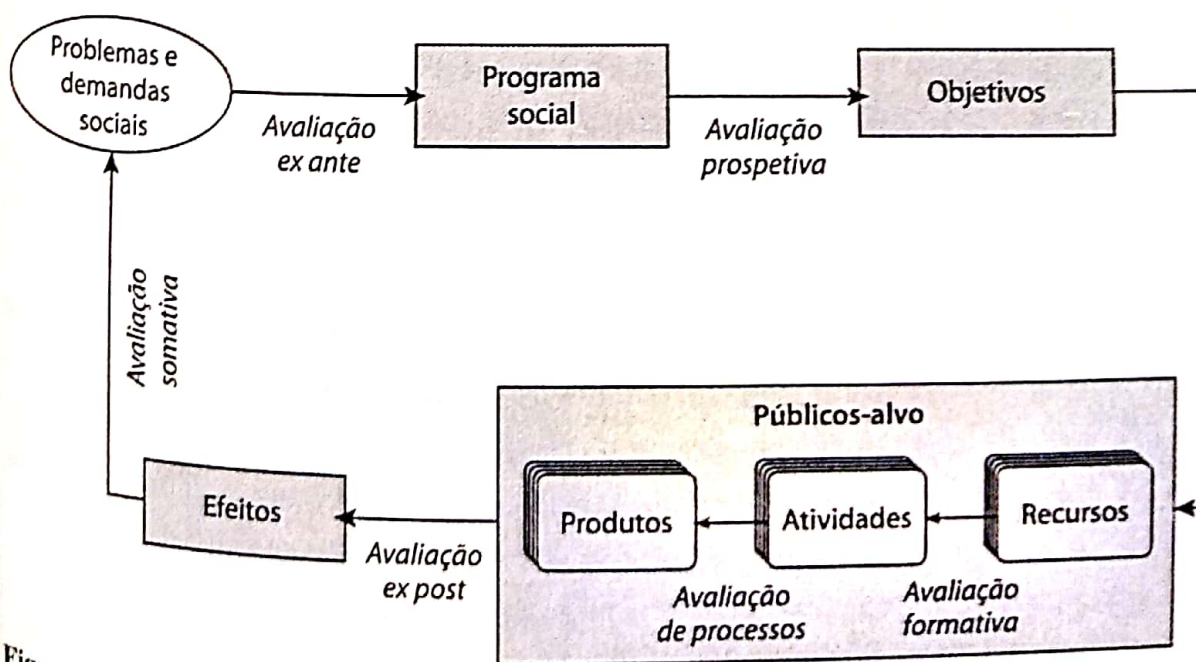


Figura 2.2. Tipos de avaliação de um Programa Social.

Cada modalidade tem vantagens e limitações. A avaliação externa parece permitir um olhar objetivo, não enviesado, não complacente com problemas recorrentes e não comprometido com a continuidade do programa, o que confere maior isenção à análise, supondo-se que os consultores são absolutamente profissionais, não preocupados com as chances de contratação em uma a próxima demanda de avaliação. Pelo envolvimento das equipes, a avaliação interna, a mista e a participativa parecem permitir aportes mais circunstanciados na análise dos processos e resultados, além de facilitar a incorporação de melhorias na gestão do programa – objetivo último da avaliação. Mas podem ser míopes aos problemas que coloquem em xeque a continuidade do programa ou tirem equipes da “zona de conforto”. Avaliações internas podem ser rápidas e baratas, ao contrário das outras modalidades. Avaliações participativas podem viabilizar aprimoramentos mais efetivos, pelo envolvimento dos beneficiários e operadores no sucesso do programa.

## Avaliação segundo o ciclo de programas

Uma tipologia de pesquisas avaliativas menos conhecida, mas particularmente interessante para organizar a avaliação de programas sociais, é a proposta por Rossi *et al.* (2004), de classificação de avaliação segundo necessidades de informação para o estágio em que se encontra o programa no seu ciclo de vida, como foi apresentado no final do capítulo anterior. Para esses autores, os estudos avaliativos deveriam compor um plano de investigação completo, seguindo a lógica implícita do ciclo de programas, começando pela análise da pertinência do programa na Agenda de Políticas Públicas – *Avaliação de Demandas Sociais* ou *Avaliação Diagnóstica*. A pergunta inicial é se o programa proposto responde a uma demanda social efetivamente existente. Trata-se de uma demanda social legítima, um compromisso genérico de campanha para o qual já existe solução implantada ou uma necessidade de preservação de uma estrutura organizacional criada em algum momento do passado, para lidar com o problema que então existia e não existe mais?

Se o programa existe para responder a uma demanda pública ainda existente de fato, então se deveria investigar a teoria de mudança social ou modelo de intervenção em que o programa se assenta – *Avaliação de Desenho*. Afinal, essa é a melhor proposta e o melhor desenho lógico de intervenção para o problema social em questão, diante de outras possibilidades de intervenção? Para promover melhoria do desempenho escolar de crianças de baixa renda em uma dada região, deve-se colocá-las em uma escola em tempo integral, diminuir o tamanho das turmas, contratar um segundo professor, mudar a didática empregada ou proporcionar melhores condições

materiais às suas famílias? O desenho proposto, de articulação federativa e envolvimento de outras áreas setoriais de governo, funciona a contento, ante as distintas capacidades de gestão e controle social pelo país afora? Os custos de programas criados para mitigar tal problemática, na escala pretendida, são sustentáveis a médio e longo prazo? Diante das questões anteriores, este programa deve ser priorizado na agenda estratégica de políticas públicas, no momento atual, com o desenho operacional proposto?

Respondidas essas questões, faria sentido, então, investigar o processo de implementação do programa – *Avaliação de Processo* – seus gargalos de entregas, os motivos de atrasos nas obras, as dificuldades na prestação de serviços com a qualidade desejável, enfim, entender os problemas críticos que afetam o programa e o cumprimento de seus objetivos. Constatada a inexistência de graves problemas de implementação, faria sentido avaliar os resultados concretos do programa para seu público-alvo e os impactos sociais mais abrangentes da intervenção – *Avaliação de Resultados e Impactos*.

Ao final, tendo-se constatado que o programa atende a uma demanda social concreta, conta com um desenho de implementação adequado, com procedimentos operacionais bem-estabelecidos e com resultados efetivamente comprovados, passar-se-ia para a *Avaliação da Eficiência* ou de *Custo-Efetividade*, para responder se o custo do programa se sustenta para escalas mais amplas de operação ou se os efeitos sociais obtidos são maiores que os de outros tipos de intervenção.

Essa proposta é certamente um avanço conceitual na forma de se entenderem os diferentes focos dos estudos avaliativos, para além daquelas dicotomias de avaliação de processos/avaliação de impactos, avaliação *ex ante*/avaliação *ex post* etc. Ao propor ênfases avaliativas para cada etapa do ciclo de vida do programa, os autores oferecem um plano de avaliação abrangente, orientando a especificação das pesquisas de avaliação segundo uma lógica de abordagem que tenderia a garantir, *a priori*, resultados relevantes e úteis ao final do processo. Tal proposta ajudaria a definir melhor a questão crucial para sucesso de uma pesquisa de avaliação de programas: que aspectos do programa devem ser avaliados e quais devem ser abordados posteriormente, em função do estágio de maturidade do programa, para maximizar as chances de se obterem informações úteis e relevantes para seu aprimoramento.

Afinal, não caberia especificar uma pesquisa de avaliação de resultados e impactos se há dúvidas com relação ao estágio de maturidade dos processos e ações previstos na implementação do programa ou se ainda existem suspeitas de que o desenho de implementação padece de supostos não sustentáveis em sua estrutura lógica de intervenção. Não caberia proceder-se a investigações complexas de *Análise de Custo-Efetividade* se não há certeza de que os resultados e impactos gerados pelo programa

são significativos. Enfim, evidências de baixa efetividade do programa poderiam ser consequências menos de falhas no seu desenho ou implementação e mais da própria inadequação de sua existência. Esse parece ser o caso dos programas voltados a oferecer experiência de primeiro emprego para jovens: embora a demanda por experiência de trabalho figure entre as prioridades de jovens, eles são mais seletivos na busca de oportunidades e mais propensos a tentar novas experiências de emprego. Não é, pois, demanda por emprego que muitos buscam, mas experiências enriquecedoras para decisões posteriores de vida profissional (Madeira, 2004). O problema a ser enfrentado não é de “primeiro emprego”, mas de “primeiras experiências e vivências no mundo do trabalho”, certamente difíceis de serem oferecidas na prática, dadas as características do mercado de trabalho brasileiro. Em outras situações, as condições de vida do jovem ou suas aspirações de consumo podem levá-lo, inclusive, a deixar a escola em segundo plano. Afinal, deveriam existir programas de “primeiro emprego”? Tornar a escola de ensino médio mais atraente para o jovem não deveria ser a prioridade, inclusive com a concessão de bolsas de estudo para alguma atividade de pesquisa, extensão ou apoio comunitário?

A hierarquia de focos avaliativos de Rossi *et al.* (2004) não implica que o gestor só possa dispor de informações acerca da eficiência, da qualidade dos serviços, da cobertura de atendimento do público-alvo nas fases mais adiantadas do ciclo de vida do programa. Pode ser necessário encomendar pesquisas de avaliação de resultados e impactos ou estudos mais aprofundados de avaliação da eficiência em estágios ainda incipientes da implementação do programa, para responder a demandas externas de avaliação. O que os autores defendem é que se organize um plano consistente de avaliação em todas as etapas do ciclo de vida do programa, e que se resista à tentação de “colocar a carroça antes dos bois”, encomendando pesquisas de avaliação de forma precoce e desarticulada. É preciso ser diligente com o risco de produção de resultados irrelevantes ou, pior, com a produção de resultados precipitados com pretensa legitimidade científica. Para um gestor de programas, mais difícil que lidar com fatos críticos acerca do programa sob sua responsabilidade é responder por factoides, construídos com base em evidências equivocadas ou mal-interpretadas em um estudo malconduzido. Avaliação é recurso de aprendizagem organizacional, não de desmobilização ou deslegitimação institucional.

## Avaliação segundo a natureza do produto de pesquisa

Outra forma menos usual de classificar as avaliações considera seu formato enquanto produto: *Relatórios-Síntese* com conjunto de tabelas e textos que resumizam indicadores de resultados de políticas e programas, para fins de divulgação a públicos mais abrangentes; *Estudos Avaliativos*, resultantes de análises com base em dados secundários, material documental ou estudos já realizados anteriormente na temática; *Meta-avaliações*, recensões sobre estudos avaliativos, pesquisas e experiências nacionais, subnacionais e internacionais de programas e projetos realizados e *Pesquisas de avaliação*, levantamentos primários, qualitativos ou quantitativos, desenhadas com objetivos de produção de evidências específicas, em geral, não disponíveis, e necessárias ao aprimoramento da intervenção.

*Relatórios-Síntese*, como os documentos, informes e publicações do IBGE, Ipea, agências multilaterais – PNUD, Unicef, Unesco, Habitat, FAO, OIT, Cepal, Banco Mundial e Banco Interamericano de Desenvolvimento – e a Iniciativa Brasileira por um Mundo Sem Pobreza (WWP) são importantes fontes de informação para avaliação de políticas, programas e ações governamentais pelo mundo (Quadro 2.1). Essas publicações, geralmente anuais, trazem análises de políticas e programas de diferentes países e diversos indicadores da realidade social. Podem-se conhecer experiências, as melhores práticas, assim como as dificuldades enfrentadas nos vários países estudados. Certamente, há que se considerar que, como toda avaliação ou estudo, a visão de mundo e os enfoques metodológicos adotados pela instituição produtora influenciam a maneira pela qual as experiências e os resultados são apresentados no relatório.

*Estudos avaliativos* com base em bibliografia ou fonte de dados secundários produzem insumos para responder a demandas de informação e conhecimento para aprimoramento de programas e, necessariamente, para especificar pesquisas empíricas de avaliação. Dissertações de Mestrado e Teses de Doutorado, apresentadas nas universidades brasileiras<sup>6</sup> e em outros países, artigos científicos publicados nas diversas revistas das Ciências Sociais Aplicadas, estudos de instituições de pesquisa constituem acervo fundamental para equipes encarregadas de avaliação de programas, ainda que o conhecimento neles registrados possam não responder às questões de investigação específicas da Administração Pública. Como todo empreendimento técnico-científico, é necessário que se faça alguma compilação/análise de estudos produzidos antes de sair a campo para coletar novas evidências empíricas acerca de determinada questão. Além disso, estudos avaliativos compreendem também análises com base em dados secundários, como as pesquisas levantadas pelo IBGE, assim

6. Vide Portal Capes com teses defendidas no país em <http://bancodeteses.capes.gov.br/>

como os Registros Administrativos e Cadastros Públicos. Há um volume considerável de dados empíricos coletados em sistemas de gestão de programas, ainda pouco explorados em termos do seu potencial para responder questões sobre efeitos integrados de políticas e programas públicos no país.

As *meta-avaliações* reúnem evidências sobre vários aspectos acerca do desenho, implementação e resultados dos programas. Ainda que se possa desenvolver um conjunto amplo de pesquisas de avaliação de programas de uma área setorial de política é preciso dispor, de tempos em tempos, de análises mais abrangentes da política setorial que os congrega. O conjunto de avaliações de programas, revisitados na forma de uma meta-avaliação, acaba por compor um quadro que permite identificar aspectos meritórios, avanços e contradições da Política Pública, trazendo insumos para sua reformulação.

As *pesquisas de avaliação* são empreendimentos avaliativos que envolvem levantamentos empíricos em campo. Tais levantamentos podem ser de natureza mais estruturada, como as *enquetes quantitativas com marco amostral probabilístico* – necessárias para produção de indicadores de dimensionamento de públicos-alvo ou inferência representativa quanto aos resultados dos programas – ou com *amostras intencionais* – mais rápidas, menos custosas, mas com limitado poder de generalização dos seus resultados. Um tipo particular de avaliação de natureza estruturada são as pesquisas com delineamento quase-experimental – também denominadas avaliação de impacto. As pesquisas podem ser, por outro lado, menos estruturadas, mais exploratórias, *de cunho qualitativo*, como grupos focais ou entrevistas em profundidade, dirigidas aos usuários e beneficiários de programas e projetos, assim como aos gestores e técnicos encarregados da operacionalização dos serviços. Na próxima seção e, especialmente, no próximo capítulo, serão discutidas, mais detalhadamente, várias metodologias empregadas comumente nas pesquisas de avaliação de programas sociais.

**Quadro 2.1.** Alguns dos relatórios sociais periódicos.

Instituição	Relatório
IBGE	Síntese de Indicadores Sociais
IPEA	Planejamento e Políticas Públicas
MDS	Cadernos de Estudos Desenvolvimento Social em Debate
RBMA	Revista Brasileira de Monitoramento e Avaliação
CEPAL	Panorama Social da América Latina e Caribe
PNUD	Relatório de Desenvolvimento Humano
UNESCO	Relatório Educação para Todos
UNICEF	Estado Mundial da Infância
FAO	Estado Mundial da Insegurança Alimentar
OIT	Tendências Mundiais do Emprego
Banco Mundial	Relatório do Desenvolvimento Mundial



Essa tipologia de avaliação enquanto formato/produto não é usualmente encontrada nos manuais da área, mas pode ser interessante para diferenciar avaliações segundo o esforço e os recursos requeridos na estruturação de uma equipe envolvida em monitoramento e avaliação na Administração Pública. A elaboração de relatórios-síntese constitui-se em produto básico que toda unidade de M&A deveria ter condições de fazer, pois requer equipes técnicas menores e com menor exigência quanto a conhecimentos especializados. Meta-avaliações, por sua vez, demandam um corpo técnico com mais experiência, mais especializado e com perfil multidisciplinar. Estudos avaliativos exigem equipe maior, com capacidade de processamento e análise estatística de base de dados ou análise mais substantiva de temas nas áreas de interesse programático. A realização ou contratação de pesquisas de avaliação pressupõem as competências requeridas para os produtos anteriores e experiência na gestão de levantamentos de campo, qualitativas ou quantitativas, definição do termo de referência e análise de dados. Neste último caso, o domínio completo do processo de especificação da avaliação de programa é crucial, tema tratado na próxima seção.

## Etapas para especificação e realização de uma pesquisa de avaliação

A realização de pesquisas de avaliação de programas pode ser dividida em seis grandes etapas de atividades: (1) estudo de avaliabilidade do programa; (2) concepção e delimitação do escopo da avaliação; (3) desenho e especificação metodológica da pesquisa; (4) condução do trabalho de pesquisa documental e/ou de campo; (5) análise de dados da pesquisa e discussão de achados com a equipe do programa; por fim, (6) a produção de relatório e disseminação de resultados.

Alguns manuais clássicos na área podem prescrever um número maior de etapas, não incluir o estudo de avaliabilidade como parte do processo formal ou, ainda, propor modelos menos lineares e unidirecionais que o retratado na Figura 2.3<sup>7</sup>. Para fins didáticos, a proposta das etapas listadas parece suficientemente robusta e justificável para ilustrar os esforços diferenciados necessários no ciclo de produção da maior parte dos estudos avaliativos de programas públicos, de menor ou maior complexidade. Evidentemente, como em todo processo de produção de conhecimento, decisões tomadas em uma etapa anterior podem ter de ser revistas, de modo que se torne mais viável a concretização das etapas posteriores. Vale citar dois exemplos que ocorrem com alguma frequência: a pesquisa de avaliação é definida com escopo

7. Vide Worthern *et al.* (2004), Rossi *et al.* (2004), Imas e Rist (2009).

abrangente, mas revela-se inviável – em custos e tempo – na etapa do seu desenho metodológico ou, definida a estratégia metodológica da pesquisa, ela se mostra, em campo, problemática para operacionalizada, por causa das dificuldades logísticas encontradas. Nesses casos, é necessário retomar a etapa anterior do planejamento da pesquisa. Nota-se, portanto, que na prática, o modelo de planejamento de pesquisas de avaliação é menos linear que o sugerido pela figura.

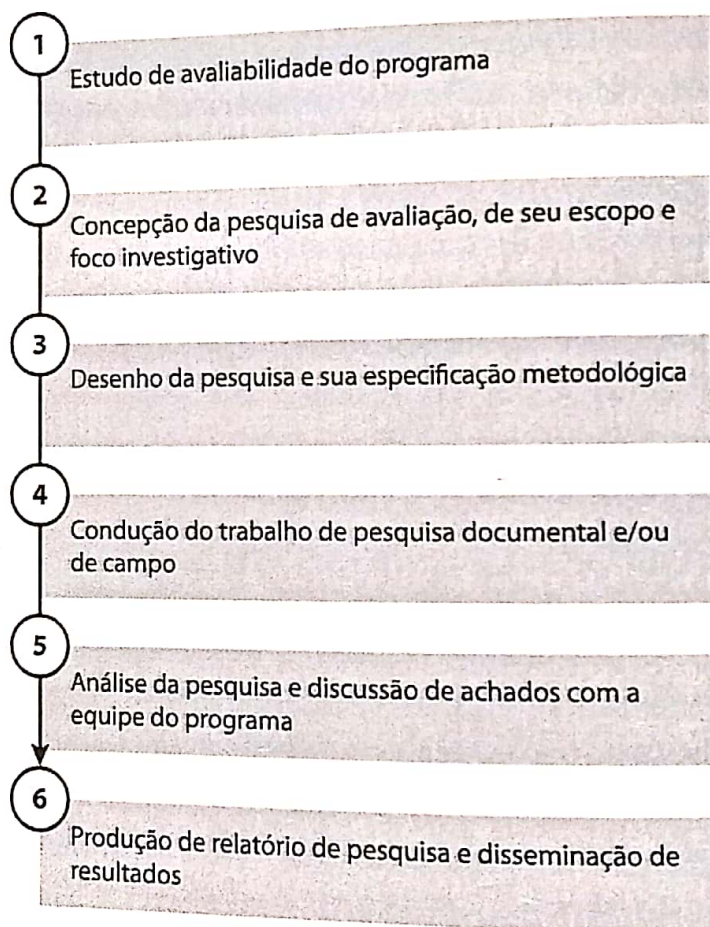


Figura 2.3. Etapas de uma pesquisa de avaliação.

O *estudo de avaliabilidade* é uma etapa crucial desse processo, já que define se a avaliação vai ser ou não realizada e que contornos deve assumir nas fases seguintes. Trata-se de analisar se ela deve ser efetuada, se pode produzir evidências relevantes para o programa em questão e se o estágio em que se encontra o programa justifica um esforço maior de investigação sobre seus componentes, o que pode ser feito por meio das seguintes questões:

Estaria o programa em estágio ou momento propício para uma investigação mais detalhada? Em que aspecto? No seu desenho de intervenção, em atividades de sua implementação ou já estaria avançado suficientemente para uma avaliação de

resultados e impactos? O programa, de fato, já existe? Houve dotação orçamentária, repasses para agentes operadores e alocação de equipes? Já está implantado no país ou região em que fora previsto? Existe um sistema de gestão, registro de atividades e indicadores de monitoramento? Existem bases de dados para avaliação do programa? Já foram feitas avaliações anteriores do programa?

Trata-se de perguntas básicas que permitem assegurar a factibilidade (ou não) de se fazer um esforço de avaliação de programa com possibilidades de produzir informações relevantes para seu aprimoramento, e qual deve ser o foco da pesquisa, considerando-se o estágio em que se encontra o programa no seu ciclo de vida.

Antecipar a avaliação de um programa que se encontra em seus primeiros estágios de implantação pode levar à produção de resultados pouco relevantes ou consistentes. Pode, também, desviar o foco e atrapalhar o processo de implantação das rotinas operacionais do programa e de resolução dos problemas básicos de qualquer novo projeto no setor público, como a obtenção dos recursos financeiros, contratação de equipes, articulação com outros setores do governo, criação de um sistema de gestão, de painel de indicadores de monitoramento, entre outros. Realizar uma avaliação de resultados e impactos sem assegurar que as rotinas básicas do programa já se encontram estruturadas e operando adequadamente pode produzir evidências de baixa ou falta de efetividade do programa. De forma apressada, pode-se deslegitimar uma boa e meritória proposta de intervenção que, por complexidade da problemática visada ou do arranjo operacional idealizado para mitigá-la, não teve seu tempo adequado de maturação. Por outro lado, postergar avaliações formativas de processos críticos pode conspirar para atraso da implantação do programa e, assim, adiar reformulações necessárias ao seu desenho original.

O estudo de avaliabilidade recupera os principais documentos descritivos e normativos do programa, sistematizando seus objetivos, públicos-alvo, lógica de intervenção, mapeamento de principais atividades e seus produtos, principais agentes públicos e privados envolvidos, tipos de serviços e entregas. Além disto, procura identificar os sistemas de gestão, bases de dados e pesquisas de avaliação já realizadas anteriormente. Certamente, requer entrevistas de técnicos e gestores, em geral, envolvidos na coordenação do programa. Idealmente, a finalização dessa etapa não deveria resultar apenas na produção de uma nota sintética com as evidências e conclusões acerca da viabilidade, pertinência e continuidade (ou não) da avaliação, mas uma documentação – síntese atualizada do programa, quem sabe, com um MaPR, recurso dialógico interessante para iniciar a discussão da próxima etapa.

Com tais características, o estudo de avaliabilidade poderia até mesmo ser considerado como um tipo de avaliação – avaliação documental ou como um tipo expedito de avaliação de desenho – ao fazer um *checklist* de aspectos básicos para a

operação do programa, trazendo insumos para equipe de coordenação acerca de eventuais lacunas encontradas na documentação descritiva e operacional que o compõe ou mesmo das próprias atividades a serem desenvolvidas por ele. Para a equipe encarregada da avaliação, esse pré-diagnóstico permite decidir-se pela continuidade ou não da pesquisa, e, nesse caso, as indicações de qual deve ser o foco avaliativo. Em situações extremas, o estudo pode mostrar que o programa não pode ser avaliado, pois não tem qualquer documentação formal, sistema de monitoramento ou base de dados organizada que possa servir de referência. Em tais circunstâncias, o estudo de avaliabilidade serve para justificar a necessidade de se proceder à documentação do programa, de seus objetivos, atividades, de redesenho do modelo de intervenção, de construção de indicadores-chave etc.

### Concepção e desenho da pesquisa de avaliação

Se o programa se revela em condições de ser avaliado, iniciam-se as atividades de concepção *delimitação dos objetivos e determinação do foco da pesquisa de avaliação*. Como já se procurou enfatizar neste livro, programas públicos são empreendimentos complexos, com muitas atividades, os quais operam em contextos de vulnerabilidade social e de capacidade de gestão muito heterogêneos. É preciso ter clareza do foco avaliativo, e, para isso, deve-se procurar definir que componentes e que aspectos do programa serão avaliados. Há tantas possibilidades de investigação em um programa social que a avaliação pode até mesmo se tornar um “programa de investigação progressivo”, no sentido lakatiano (Lakatos, 1979), o que certamente é um efeito interessante para orientar pesquisas acadêmicas, mas não os esforços pragmáticos na pesquisa de avaliação no setor público, em que tempestividade e objetividade são critérios cruciais na decisão do que avaliar e de como fazê-lo.

É importante que a etapa em que o programa se encontra já aponte as demandas centrais de avaliação, como é ilustrado na apresentação da tipologia de Rossi *et al.* (2004) na seção anterior. Se o programa foi recém-implantado, há sempre necessidade de avaliar processos e atividades, identificar gargalos e a qualidade dos serviços prestados para os públicos-alvo. Mas, se já está operando há muito tempo, pode ser o caso de uma avaliação somativa, de análise das contribuições do programa na mitigação do problema que o originou. Entretanto, mesmo avaliações formativas ou somativas podem abordar muitos aspectos da implementação ou efeitos (positivos ou negativos) do programa, por exemplo: por que determinadas atividades não estão se desenvolvendo plenamente? Por que, em algumas regiões, o programa obtém mais resultados? Os serviços estão sendo prestados satisfatoriamente, de acordo com padrões de qualidade definidos? Qual a cobertura do programa? Quem está se beneficiando dele?

A equipe coordenadora do programa pode já ter algumas questões de interesse, às vezes muito específicas, de natureza formativa. O decisor estratégico, a quem cabe negociar recursos para o programa, pode ter outras demandas, do tipo de avaliações *ex post*, visando colher argumentos técnicos para mais recursos. Diferentes equipes e gestores podem ter visões distintas do que deve ser avaliado e a especificação do objetivo da avaliação pode levar mais tempo que se supõe. Reuniões técnicas e dinâmicas de grupo com circulação prévia de relatórios parciais ajudam formar consensos e clarificar as questões centrais a investigar.

O MaPR pode ser um recurso para facilitar a discussão com os interessados acerca dos diferentes focos de maior interesse avaliativo. Como já foi apresentado no capítulo anterior, o mapa provê uma narrativa do programa, explicitando, de forma mais geral, como seus objetivos são “produzidos” pelas atividades, previstas segundo uma certa lógica de articulação, pelo contexto e pelas condições necessárias para que se desenvolvam. Pode ser detalhado em subprocessos, no componente de maior interesse. É, portanto, um mapa que pode levar a outro mapeamento: o das pesquisas de avaliação a serem realizadas durante um período mais longo, como é ilustrado, mais à frente, neste capítulo.

Em dado momento, até por necessidade de finalização do esforço de concepção, emerge o tema de maior interesse, a relevância estratégica ou criticidade, a ser analisado no programa. É o início do *desenho metodológico* da avaliação. Nessa etapa, a terceira do processo de planejamento da avaliação, são formuladas as perguntas gerais a serem respondidas, os contextos territoriais ou arquetípicos de interesse e os sujeitos a serem abordados – gestores estaduais, técnicos municipais, beneficiários do programa ou usuários dos serviços, a população em geral. É necessário escolher a técnica de coleta mais adequada para responder às questões levantadas, a natureza e tamanho da amostra de sujeitos ou instituições a serem entrevistadas, as questões éticas suscitadas, considerando os prazos e recursos disponíveis.

Quanto às técnicas, a pesquisa social oferece várias alternativas: a compilação de estudos anteriormente já realizados sobre questões similares, a busca de dados secundários, a observação e suas variantes – visitação, observação etnográfica etc. –, as entrevistas – semiestruturadas ou estruturadas –, os grupos de discussão, as pesquisas de campo com questionários – com amostras probabilísticas ou intencionais –, e os desenhos quase-experimentais. Esses métodos podem ser classificados em dois grandes campos, de acordo com o tipo de perguntas e estruturação dos instrumentos de coleta e análise das informações para respondê-las: métodos qualitativos e métodos quantitativos. Simplificadamente, pode-se dizer que os métodos qualitativos estão voltados para um entendimento, mais exploratório ou em profundidade, da problemática social em questão; já os métodos quantitativos são mais adequados para captação

de aspectos mais tangíveis da realidade social. No primeiro caso, é preciso especificar um roteiro de aspectos a observar ou de questões a perguntar; no segundo, formular um questionário com perguntas estruturadas e respostas, em geral, pré-codificadas.

Não cabe aqui adentrar-se a vastidão dos métodos e técnicas da pesquisa social. No capítulo seguinte, faz-se uma breve sistematização daqueles mais empregados na avaliação de programas, assim como também se faz uma discussão sobre definição da amostra – probabilísticas ou intencionais – em pesquisas qualitativas e quantitativas, questões éticas do desenho e condução de pesquisas empíricas e sobre a conexão causal entre o programa e seus possíveis impactos. Vale adiantar que não se advoga por nenhum método ou técnica em particular na avaliação de programas. Ao contrário do que defendem determinadas comunidades de avaliadores, não existe um padrão-ouro de pesquisa avaliativa. Neste livro, em consonância com comunidades mais ecléticas e multidisciplinares de avaliação, defende-se que a melhor técnica de investigação é aquela mais apropriada às perguntas, objetivos, contextos, prazos e recursos disponíveis para avaliação.

Da finalização do processo de especificação da metodologia da pesquisa de avaliação ao início do trabalho de campo da pesquisa, é preciso definir como a pesquisa será efetivamente realizada: pela equipe interna do programa, com a contratação de equipe externa – instituto de pesquisa, firma de consultoria, consultores especializados ou fundação universitária –, ou por meio de um arranjo misto dos listados anteriormente. Contratações externas de pesquisas de avaliação, no setor público, envolvem esforços expressivos de definição dos termos de referência para contratação, para garantir que, ao final do processo, o relatório traga, de fato, os insumos esperados, relevantes e consistentes para sua apropriação junto às equipes. Pelos cuidados e riscos descritos por Pinto *et al.* (2013), a elaboração de termos de referência para contratação de pesquisas de avaliação para programas sociais parece ser, mais que uma técnica, uma “disciplina especializada” do Direito Administrativo.

Um dos elementos centrais na contratação externa são os gastos da pesquisa. Custos com equipe de campo, supervisão, consistência e análise precisam ser orçados não apenas pelo tamanho, mas, também, em levantamentos para avaliação de programas, com base na qualificação técnica e experiência exigida. Pesquisas de avaliação não podem ter como referência de custos os parâmetros usados em enquetes de opinião. Estas últimas têm roteiros ou questionários muito mais simples, empregam entrevistadores com menor formação escolar, valem-se de amostras não probabilísticas (cotas etc.). São desenhadas para prover respostas rápidas para problemas mais simples, investigados de forma um tanto recorrente. Trata-se de um contexto bastante distinto das pesquisas de avaliação de programas, como se tem mostrado nestes primeiros capítulos do livro.

As últimas atividades do desenho metodológico da pesquisa envolvem a construção do roteiro básico de observação/entrevista no caso de abordagens qualitativas de coleta de dados ou do questionário, quando de levantamentos mais estruturados, com amostras mais amplas. Em um caso ou outro, é preciso ter clareza de que o que se quer investigar não é diretamente o que se vai observar ou perguntar aos sujeitos da amostra. Se, por exemplo, há o interesse em investigar o preconceito de raça ou da condição social de uma pessoa, parece pouco efetivo questionar diretamente à pessoa se ela “tem preconceito ou discrimina alguém” por tal ou qual motivo. Uma bateria de questões mais gerais – você concorda que as pessoas nascem com direitos iguais etc. –, com escala de concordância/discordância, seria um caminho melhor. Se há o interesse, por exemplo, de avaliar a capacidade de gestão de programas em uma unidade de prestação de serviços ou município, há que se criar um protocolo detalhado de levantamento de informações, com dados sobre recursos humanos, equipamentos etc. Se há interesse, por outro lado, de avaliar a qualidade de um serviço público, não basta pedir ao entrevistado que o avalie com uma nota (em um intervalo). É preciso saber que aspectos dos serviços prestados não estão a contento. Enfim, a elaboração do roteiro ou questionário abrange muita reflexão para traduzir as demandas de investigação em questões operacionalmente factíveis em campo e que, na análise, contribuam para construir a narrativa interpretativa da avaliação do programa.

### **Do trabalho de campo à produção dos resultados**

O *trabalho de campo* é a quarta etapa, também crítica, do processo de planejamento da avaliação. Pela natureza operacional das atividades envolvidas pode não gozar do *status* que se atribui às outras etapas “mais nobres” da avaliação, desenvolvidas em escritório, mas isso não diminui, de modo algum, a importância dessa fase e a necessidade de rigor técnico e zelo na sua condução. Todo esforço de concepção e especificação metodológica pode ser colocado a perder se não forem observados os cuidados necessários com o treinamento de equipe, pré-teste de instrumentos, planejamento logístico da coleta, supervisão de campo e análise crítica das evidências coletadas na pesquisa. Fazer estimativas conservadoras de tempo e recursos para essas atividades pode ser um risco.

Se a avaliação não implica extenso trabalho de campo, mas se constitui em um estudo avaliativo baseado na revisão de outros textos já produzidos ou análise de dados secundários, é recomendável também ter um acompanhamento muito próximo e periódico, para que o técnico ou pesquisador envolvido não enverede por questões secundárias, às vezes de seu maior interesse, esquecendo-se de atender as demandas avaliativas que orientaram a especificação do estudo.

Pesquisas qualitativas têm, em geral, amostras com tamanhos menores, mas requerem entrevistadores de campo mais experientes, com conhecimento sobre a temática investigada, com capacidade de improvisação criativa e adaptação. Entrevistas em profundidade ou grupos de discussão podem desconcertar um entrevistador ou moderador pouco experiente. Levantamentos quantitativos demandam esforços de coleta em amostras mais amplas, o que exige entrevistadores com prática de abordagem pessoal, capacidade de localização e disponibilidade para deslocamento pelo território. Em um caso e noutro, a equipe de campo contratada precisa estar comprometida com os objetivos da pesquisa, com a qualidade do trabalho e bem treinada nos conceitos e quesitos do instrumento de coleta.

O treinamento para coleta não pode ser subestimado em tempo e profundidade. É necessário apresentar os objetivos da pesquisa, a justificativa de sua realização, os conceitos básicos subjacentes aos quesitos investigados e fazer ensaios práticos de coleta. A duração depende da qualificação do pessoal e complexidade do questionário. Fazer ensaios de entrevistas, chamar a atenção para problemas mais frequentes nos levantamentos, remeter a equipe à leitura do manual de conceitos básicos e procedimentos operacionais são estratégias recomendadas em treinamentos com equipes grandes de entrevistadores.

O roteiro de uma pesquisa qualitativa ou o questionário de uma pesquisa quantitativa precisa ter sido testado anteriormente (pré-teste), para avaliar tempo de entrevista, assim como as melhores estratégias de coleta de determinadas informações e para reformular ou retirar perguntas que se revelem inadequadas ou inócuas. A logística de campo é outro desafio, de compatibilização de tempo de coleta e custos. Equipes maiores de entrevistadores de pesquisa podem encurtar o tempo de coleta, mas ao custo bem maior de treinamento e supervisão de campo. Equipes menores, com técnicos qualificados e bem-remunerados, podem produzir um trabalho de campo mais consistente e homogêneo.

A supervisão engloba a coordenação da equipe de campo pelo território, orientação quanto às dúvidas no preenchimento do questionário e outras dificuldades inevitáveis que surgem – por mais que se antecipem os possíveis problemas – e verificação da realização das entrevistas com base em algum critério que torne isso viável – porcentagem da amostra, escolha de locais de acesso mais difícil etc. Uma boa equipe de supervisão pode equacionar em tempo situações de mapas desatualizados de roteiros de pesquisa, de elevada taxa de recusas e de domicílios fechados, de fraudes e vieses de aplicação equivocada de questões decorrentes de problemas de treinamento. Integram o conjunto de atividades relacionadas à produção da pesquisa os procedimentos de crítica ou análise de consistência dos dados coletados em campo, para avaliar se existem desvios sistemáticos na aplicação de determinados quesitos



ou regiões da amostra planejada. Há variáveis ou quesitos mais sujeitos a problemas na coleta, como rendimentos ou recebimento de benefícios sociais, que dependem menos da destreza do entrevistador e mais da postura do entrevistado. Nesse caso, podem derivar do constrangimento do entrevistado ou de sua dificuldade de ter uma referência regular do valor de sua remuneração para informar.

Informações sobre ocorrências de um passado mais remoto igualmente são problemáticas. A crítica deve conter também análises sofisticadas de cruzamento de dados coletados em diferentes quesitos, o que implica a disponibilidade de equipe técnica dimensionada e qualificada para a tarefa. A agilidade com que se consegue diagnosticar eventuais problemas é decisiva para que a supervisão da pesquisa possa proceder ajustes e fazer orientações em tempo de garantir a coleta de dados empíricos com a qualidade requerida para a avaliação. A análise da consistência dos dados ou entrevistas estende-se, em geral, para além do período previsto de coleta de campo, o que implica, eventualmente, a volta aos sujeitos ou instituições entrevistadas para checagem ou correção da informação. Ao final do trabalho de campo, espera-se poder dispor de uma base de dados (em pesquisas quantitativas) ou registros transcritos de entrevistas, observações ou grupos focais devidamente consistidos.

O trabalho de campo não precisa ter terminado para se iniciar a *análise dos dados e interpretação dos resultados* da pesquisa, penúltima fase do planejamento da avaliação de programas. A exploração inicial dos dados coletados com a pesquisa ainda em campo pode, inclusive, ajudar a identificar eventuais problemas de coleta e corrigi-los, como foi comentado há pouco. Ademais, tal exploração permite ir apreciando alguns achados e testando algumas das hipóteses aventadas quando da concepção da pesquisa de avaliação, e adianta tarefas de análise cujo tempo pode acabar sendo encurtado pelos atrasos da coleta de campo e pela pressão das datas críticas de apresentação de resultados para equipe de coordenação do programa.

Idealmente, antes da análise substantiva dos achados da pesquisa, é boa prática fazer nova avaliação da consistência dos dados de campo, agora com a perspectiva da comparação de variáveis ou indicadores da pesquisa ao longo do tempo ou de territórios, se for possível comparar com outros levantamentos. Também denominada de apuração prévia dos resultados, essa atividade pode constituir, de fato, uma etapa particular da pesquisa, além das seis já listadas, situação concreta dos grandes levantamentos amostrais realizados por instituições de pesquisa como IBGE.

Os resultados dessa avaliação podem apontar para a necessidade de ajustes, imputação de valores em variáveis sem resposta ou com respostas inconsistentes, reponderação da amostra, reclassificação de variáveis e outros procedimentos que, baseados em métodos bem-estabelecidos e transparentes, são defensáveis e aprimoram a qualidade da informação coletada em campo. Naturalmente, há situações em

que a informação padece de vícios ou vieses de coleta insanáveis, em que seu descarte é melhor decisão técnica a tomar. Há casos, pois, em que se aplica a máxima de que “Pior do que não ter uma informação, é dispor de uma informação enviesada que, com pretensão rigor técnico, pode orientar uma decisão equivocada”.

Tal como para coleta de informações em campo, há diversos métodos de pesquisa para auxiliar na análise dos resultados da pesquisa, ainda que as técnicas empregadas na coleta determinem, em boa medida, as abordagens analíticas que podem ser adotadas. Pesquisas qualitativas remetem a abordagens mais ensaísticas e narrativas, articuladas com base na análise de conteúdo dos relatos e entrevistas ou, ainda, a técnicas mais estruturadas, como a análise estruturada de textos. Dados de pesquisas quantitativas podem ser analisados de uma perspectiva mais descritiva, por meio da criação de categorias de agregação das informações, tabelas cruzadas e técnicas exploratórias, ou por intermédio de métodos estatísticos inferenciais, menos ou mais complexos, com menor ou maior exigência de supostos quanto às propriedades empíricas dos dados coletados. Qualquer que seja a abordagem ou técnica, é necessário ter um modelo teórico – eclético tanto quanto possível diante da complexidade do objeto de estudo – ou um plano analítico mais pragmático – relacionado às perguntas motivadoras da avaliação – para orientar as escolhas metodológicas e exploração dos dados empíricos levantados. Modelos teóricos multidisciplinares das Ciências Sociais ou perspectivas mais pragmáticas explicitadas em um MaPR do programa avaliado auxiliam na construção do plano tabular – ou modelo estatístico – em pesquisas quantitativas ou na elaboração das narrativas – ou mapas conceituais estruturados – em pesquisas qualitativas.

Certamente, não é a sofisticação da técnica analítica que confere legitimidade científica à análise ou relevância substantiva à pesquisa de avaliação. Não é o emprego de um modelo estatístico, econométrico ou análise quantitativa de dados qualitativos que pode garantir aportes relevantes e robustos da pesquisa.

Há certa mitificação – senão, mistificação – das abordagens ‘quantitativistas’ complexas na pesquisa acadêmica em disciplinas das Ciências Sociais Aplicadas, que acabam influenciando as pesquisas de avaliação. É como se a complexidade do modelo quantitativo pudesse substituir a “imaginação sociológica” do pesquisador na busca das respostas para as perguntas de interesse ou como se o emprego de tais técnicas pudesse conferir neutralidade na interpretação dos dados de pesquisas de avaliação.

Discurso científico, da perspectiva kuhniana, é antes uma narrativa plausível de acontecimentos e encadeamentos, às vezes, inesperados, mas mais verossímil para a comunidade epistêmica envolvida que uma pretensa elucidação neutra e insuspeita, como gostariam os positivistas, ainda tão presentes na comunidade de avaliadores e de pesquisa acadêmica (em que pese a crítica quase centenária de tantos filósofos e sociólogos da Ciência) (Kuhn, 1992).

Como foi comentado, alguns parágrafos atrás, quanto aos métodos de coleta, também não é possível ou factível adentrar no universo de métodos de análise de dados de pesquisas nesta obra, ainda muito mais vasto que o de técnicas de coleta. Com o avanço da análise dos dados e maturação interpretativa dos resultados, inicia-se a sexta e última etapa do processo de planejamento da avaliação: *a produção do relatório final e definição das estratégias de disseminação* dos resultados da pesquisa de avaliação. Idealmente, os resultados da pesquisa devem ser compartilhados, se consistentes, com as equipes da pesquisa e do programa envolvidas na avaliação, desde o início da análise. Trata-se de uma boa estratégia para, gradualmente, preparar a equipe do programa acerca dos resultados mais incisivos e negativos que podem vir ao final, e receber contribuições para interpretação dos resultados.

Como advogam Imas e Rist (2009), o relatório de pesquisa deve ser claro, organizado por temas gerais ou de acordo com as grandes perguntas de interesse da avaliação. Sua finalidade é facilitar que os resultados e os destaques sejam entendidos pela equipe do programa, seus operadores e seus formuladores. O objetivo do relatório deve ser o de comunicar os achados aos seus principais interessados e não o de impressionar a comunidade de avaliadores pela complexidade das técnicas ou ineditismo da abordagem metodológica adotada, critérios que presidem, muitas vezes, a elaboração de relatórios de pesquisa. É preciso também não confundi-lo com um trabalho elaborado para submissão a uma revista científica, em que a organização dos tópicos reflete uma lógica que pode não ser a melhor para o público de técnicos, gestores e decisores que se quer atingir. Também não se deve pretender exaustivo na apresentação dos resultados. Organizar os resultados segundo relatórios temáticos pode ser uma boa estratégia, sobretudo nas avaliações somativas, que abordam muitas problemáticas diferentes, em grande profundidade analítica. Não adianta imaginar que todos os interessados na avaliação dispõem de tempo para ler relatórios exaustivos, escritos sem preocupação com a inteligibilidade para não especializados nas técnicas empregadas na pesquisa.

Assim, não é possível contemplar todos os interesses por meio de apenas uma publicação ou estratégia de comunicação. É preciso ter várias estratégias de disseminação: resumos por e-mail, relatórios, sumários, videoconferências, apresentações. Produtos eletrônicos podem ser elaborados com uso de *hiperlinks*, mantendo, no corpo principal, apenas os aspectos centrais da avaliação. Tabelas, gráficos e seções metodológicas podem, assim, ter uma forma de disseminação que garanta alguma possibilidade de consulta. O uso de gráficos e diagramas – desde que não complexos – são recursos interessantes para se empregar. Boas fotografias comunicam mais que muitos parágrafos ou narrativas em apresentações orais. Os sumários executivos não podem ser extensos e não replicam necessariamente o capítulo de conclusões do

relatório, pois são mais sintéticos e devem recuperar os objetivos, descrição sumária da estratégia metodológica e, principalmente, os achados que respondem às perguntas principais, formuladas na etapa de concepção da pesquisa de avaliação.

O entendimento de pesquisas de avaliação como instrumentos de aprendizagem organizacional e de aporte de conhecimentos para inovação de programas requer esforços intensos e criativos para disseminação de seus resultados, para todos os seus potenciais usuários, dos “decisores” estratégicos aos operadores dos programas na ponta, em contato com os usuários e beneficiários. É necessário também ter produtos de divulgação para sociedade, fiadora e financiadora, em última instância, dos programas de pesquisas de avaliação. Os produtos de disseminação de resultados precisam estar, portanto, ajustados à linguagem dos vários públicos visados e envolver diferentes meios e estratégias. Notas em boletins de circulação entre operadores de programas, participação em eventos com técnicos de programas, em congressos científicos, por exemplo, são algumas dessas iniciativas a serem planejadas. No quinto capítulo, será retomada a importância das estratégias de disseminação como um dos componentes estruturantes de unidade de monitoramento e avaliação em políticas e programas públicos.

Realizar uma avaliação envolve esforços vigorosos de mobilização de recursos e conhecimentos de naturezas diversas. Não é para menos, em se tratando da busca de respostas para entendimento da complexidade das problemáticas sociais e das intervenções programáticas idealizadas para mitigá-las. Para ilustrar tal diversidade de combinações de tipos e focos de pesquisas de avaliação, concebidas para as diferentes demandas de interesse para aprimoramento de programas sociais, apresentam-se, na seção seguinte, vários estudos possíveis para o fictício programa de qualificação profissional no setor de turismo.

## Mapa de processos e resultados na especificação de uma pesquisa de avaliação

Como foi comentado durante a exposição das etapas de especificação de uma pesquisa de avaliação, o MaPR pode ser um recurso valioso para identificar e pactuar o conjunto de estudos necessários para a produção de informação e conhecimento sobre um programa, em todas suas fases, no seu ‘ciclo de vida’. Ao apresentar de forma esquemática seus principais processos, da alocação de recursos, execução de atividades, entrega de produtos e produção de resultados e impactos, além do contexto e os condicionantes para sua plena operação, o MaPR evidencia mais claramente as demandas de estudos e pesquisas necessários para garantir um desenho adequado para

a intervenção programática. Ajuda, inclusive, a justificar escolhas dos esforços de avaliação e interpretar os achados nas pesquisas, vinculando-os a uma narrativa mais completa e articulada com a complexidade lógica do programa.

Valendo-se do MaPR do programa de qualificação profissional em uma região pobre, com potencial turístico, apresentado no capítulo anterior, é possível ilustrar várias das demandas avaliativas possíveis, dependendo do estágio de implantação do programa e/ou tipo de agente demandante (Figura 2.4, p. 71). Certamente, para garantir a adesão de prefeitos e inserir tal programa na agenda de prioridades de políticas de desenvolvimento dos municípios na região, foi necessário realizar um estudo prospectivo anterior, de identificação de projetos estratégicos de desenvolvimento regional, com base em registros secundários, entrevistas com agentes e lideranças locais e audiências públicas acerca de atividades econômicas, de infraestrutura de rodovias e características geográfico-ambientais.

Tendo-se identificado o potencial turístico como possível vocação regional, é necessário realizar um mapeamento mais detalhado para esse setor econômico, com análise dos registros existentes de atividades afins na região como hotelaria, artesanato, comércio, alimentação, serviços etc. Se identificada a necessidade de formação de mão de obra especializada nessas áreas, é preciso compilar as avaliações já existentes sobre programas de qualificação profissional na região, estado ou país, de forma que se possibilite o dimensionamento de recursos orçamentários e institucionais necessários para implantar um programa dessa natureza na região. O diagnóstico das condições de vida da população, das características da mão de obra e do mercado de trabalho e a demanda por turismo também compõem o espectro de possíveis pesquisas a realizar. Estas são pesquisas de avaliação do tipo *ex ante*, de diagnóstico de demanda e de desenho de programas ou, ainda, prospectivas, típicas da primeira etapa de concepção e implantação de um programa, que fornecem parâmetros para análise da viabilidade deste e de sua própria avaliabilidade.

Se o programa vai ser, de fato, iniciado, diversos outros levantamentos são necessários, mas nem tudo será realizado nos momentos inicialmente planejados em função da pressão por começar a produzir “entregas”. O universo das pesquisas de avaliação é distinto do microcosmo dos estudos acadêmicos: os tempos e movimentos das primeiras não estão na governança do coordenador da avaliação como no líder coordenador do programa de investigação acadêmico. Mas, idealmente, será efetuado, com maior ou menor fôlego, um mapeamento de instituições que podem oferecer programas de qualificação profissional na região, assim como de outros serviços associados (intermediação de mão de obra, oferta de microcrédito, assistência técnica, formalização de negócios etc.). Avaliação da infraestrutura física, das salas de aula, proximidade aos meios de transportes e acessibilidade às pessoas com deficiências é

outro tipo de pesquisa necessária, orientada, no mínimo, por meio de roteiros estruturados de observação.

Para especificação do elenco de cursos é necessária a elaboração de um estudo de demanda de mão de obra, apoiado em registros do mercado formal de trabalho (Ráis e Caged), assim como em levantamentos qualitativos por meio de grupos de discussão, dinâmicas de grupo ou audiências públicas com diversos atores: técnicos municipais de Secretarias da Educação, Assistência Social, Trabalho e Desenvolvimento Econômico; lideranças políticas, empresariais e sindicais. Outra pesquisa de avaliação requerida para preparação do material didático e adequação de linguagem e didática dos professores é um levantamento amostral para caracterização do perfil educacional da população, de sua trajetória ocupacional, experiências com qualificação e os atributos potenciais para as atividades de turismo (produção de artesanato, atendimento ao público, serviços de camareira etc.). Dimensionar a carga horária de trabalho ou estudo do público potencial e a sua disponibilidade de turno e período para realização dos cursos é outra informação básica para garantir a adesão aos cursos pelos alunos e a sua permanência neles.

Com o início das ofertas dos cursos, outras demandas avaliativas passam a orientar as pesquisas, de forma que se colem informações sobre o desempenho das atividades básicas e eventuais insumos para correção de problemas identificados. Pesquisas de avaliação da implementação começam a ser realizadas para análise da cobertura dos públicos-alvo atendidos nos cursos (se residentes nos municípios do consórcio implementador do programa, se representativos da diversidade socioeconômica regional, se jovens, mulheres ou adultos estão participando do programa etc.) e para avaliação do nível de conclusão/evasão e aprovação/reprovação de cada curso, pelas diferentes escolas ofertantes (assim como a motivação da eventual desistência). Se existe sistema de gestão escolar, parte dessas demandas pode ser atendida por indicadores de monitoramento, como se discutirá em capítulo mais à frente. Também pode ser importante fazer a avaliação da satisfação dos alunos quanto ao conteúdo dos cursos, didática do professor e material didático, por meio de estratégia multimétodos (questionários em sala de aula, observação, grupos de discussão com egressos etc.), com amostras intencionalmente escolhidas (por critérios orientados pelos níveis de evasão ou reprovação, turnos matutino e noturno etc.).

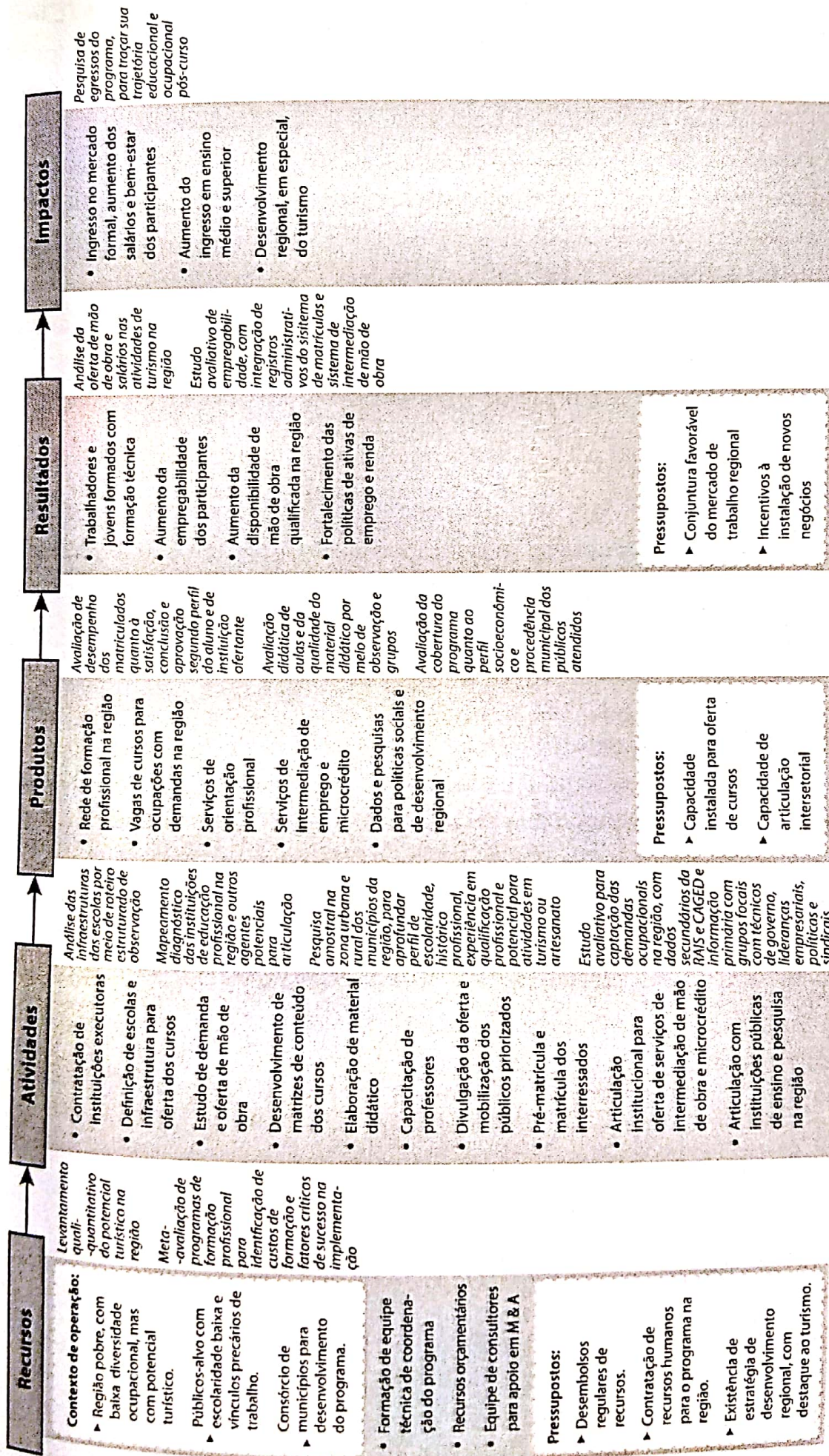


Figura 2.4. Pesquisas de Avaliação de um Programa de Qualificação Profissional em uma região pobre, com potencial turístico.

Resolvidos os problemas de implementação do programa, é natural que o foco avaliativo passe a ser as avaliações *ex post*, centradas em resultados e impactos do programa. Uma dessas avaliações pode ser concebida por meio de pesquisas amostrais com egressos, para identificar os desdobramentos dos cursos em termos de trajetória escolar (volta à escola para conclusão da educação básica, ensino médio, formação profissional complementar ou ingresso em curso superior) e a mobilidade ocupacional (ocupações em serviços relacionados ao turismo, aumento de salários ou formalização do vínculo). Pode-se também voltar aos mesmos agentes públicos e privados entrevistados nas fases iniciais do programa para avaliação do atendimento da demanda por mão de obra nas áreas afins do desenvolvimento turístico, seja em termos quantitativos, seja em termos das competências e habilidades desenvolvidas pelos cursantes. Na expectativa de que os investimentos para desenvolvimento da infraestrutura turística tenham sido realizados e o fluxo de visitantes tenha aumentado, é oportuno e relevante avaliar a empregabilidade proporcionada aos egressos, considerando-se a maior ou menor sinergia com outras ações integradas ao plano (serviços de intermediação de mão de obra, acesso ao microcrédito etc.) e valendo-se de registros de atendimentos nos sistemas de gestão do programa e dos serviços associados.

Cada uma das pesquisas aqui apresentadas tem um desenho metodológico específico, que combina técnica de coleta, sujeitos, tipo de amostra e método de análise ajustados às perguntas do que se quer avaliar. No capítulo seguinte, discorre-se sobre alguns métodos e técnicas para usados na etapa de planejamento e condução das pesquisas de avaliação de programas sociais.