

---

## CAPÍTULO 1

# Por que Avaliar?

Os programas e políticas de desenvolvimento são, normalmente, elaborados para alterar resultados como, por exemplo, aumentar a renda, melhorar o aprendizado ou reduzir doenças. Se estas mudanças são, realmente, alcançadas ou não é uma questão crucial de política pública, mas que não é examinada com frequência. Gestores de programas e formuladores de políticas concentram-se, mais comumente, em controlar e medir os insumos e produtos imediatos de um programa, tais como o total de gastos ou o número de livros escolares distribuídos, ao invés de avaliar se os programas atingiram os objetivos pretendidos de melhoria do bem-estar.

## Formulação de Políticas com Base em Evidências

As avaliações de impacto fazem parte de uma agenda mais ampla: a da formulação de políticas com base em evidências. Essa crescente tendência global está marcada por uma mudança no enfoque, de insumos para resultados. Dos Objetivos de Desenvolvimento do Milênio, passando pelos incentivos por desempenho que são pagos a prestadores de serviços públicos, essa tendência global está remodelando a forma de execução das políticas públicas. O foco nos resultados está sendo usado não somente para definir e monitorar metas nacionais e internacionais; de forma crescente, vem também sendo usado e exigido por gestores de programas para aperfeiçoar a

prestação de contas, informar a alocação orçamentária e orientar as decisões relacionadas a políticas.

O monitoramento e a avaliação estão no cerne da formulação de políticas com base em evidências. Fornecem um núcleo básico de ferramentas que as partes interessadas podem usar para verificar e melhorar a qualidade, eficiência e efetividade das intervenções nas várias etapas de execução ou, em outras palavras, focar em resultados. Podem ser encontrados agentes que fazem uso do monitoramento e da avaliação tanto em organismos governamentais quanto fora deles. Em agências governamentais ou ministérios, os agentes públicos geralmente precisam argumentar e defender o funcionamento de determinado programa junto a seus superiores e, assim, obter recursos para continuá-lo ou expandi-lo. Em nível nacional, os ministérios concorrem uns com os outros por recursos do Ministério da Fazenda. E, finalmente, os governos como um todo também têm interesse em convencer seus eleitores de que os investimentos realizados têm retornos positivos. Neste sentido, a informação e as evidências se tornam meios para facilitar a conscientização do público e promover a responsabilidade governamental. As informações produzidas pelos sistemas de monitoramento e avaliação podem ser divulgadas regularmente aos cidadãos, para informar-lhes sobre o desempenho de programas governamentais e construir uma base sólida de transparência e prestação de contas.

Em um contexto no qual os formuladores de políticas e a sociedade civil exigem resultados e a cobram prestação de contas dos programas públicos, a avaliação de impacto pode oferecer evidências críveis e robustas quanto ao desempenho e, fundamentalmente, quanto a se um programa específico atingiu os resultados desejados. Em nível mundial, as avaliações de impacto são centrais à construção do conhecimento sobre a efetividade de programas de desenvolvimento, esclarecendo o que funciona e o que não funciona na diminuição da pobreza e na promoção do bem-estar.

Simplificando, uma avaliação de impacto avalia as mudanças no bem-estar dos indivíduos que podem ser *atribuídas* a um projeto, programa ou política em particular. Este enfoque na atribuição do resultado é o selo distintivo das avaliações de impacto. Igualmente, o desafio central da execução de avaliações de impacto é identificar a *relação causal* entre o projeto, programa ou política e os resultados de interesse.

Conforme será discutido adiante, as avaliações de impacto geralmente estimam os impactos médios de um programa sobre o bem-estar dos beneficiados. Por exemplo, a introdução de um novo currículo aumentou a pontuação nos testes dos estudantes? Um programa de água e saneamento aumentou o acesso à água tratada e melhorou os resultados de saúde? Um programa de treinamento de jovens foi efetivo em estimular o

empreendedorismo e o aumento da renda? Além disso, se uma avaliação de impacto inclui uma amostra suficientemente grande de beneficiários, os resultados também podem ser comparados entre subgrupos de beneficiários. Por exemplo, a introdução de um novo currículo aumentou a pontuação nos testes entre estudantes do sexo masculino e estudantes do sexo feminino? As avaliações de impacto também podem ser usadas para testar explicitamente as opções de programas alternativos. Por exemplo, uma avaliação pode comparar o desempenho de um programa de treinamento versus o desempenho de uma campanha promocional para desenvolver a educação financeira. Em cada um destes casos, a avaliação de impacto fornece informações sobre o impacto geral de um programa, em contraste com estudos de casos específicos, que podem fornecer apenas informações parciais e podem não ser representativos dos impactos gerais de determinado programa. Neste sentido, avaliações bem formuladas e bem executadas podem fornecer evidências abrangentes e convincentes, que podem ser usadas para informar as decisões sobre políticas e formar a opinião pública. O resumo no Quadro 1.1 ilustra como a avaliação de impacto contribuiu para o debate sobre a expansão de um programa de transferência condicionada de renda no México<sup>1</sup>. O Quadro 1.2 ilustra como a avaliação de impacto ajudou a melhorar a alocação de recursos do governo da Indonésia, documentando quais políticas foram mais efetivas na diminuição das taxas de fertilidade.

## O que é Avaliação de Impacto?

A avaliação de impacto figura entre uma série abrangente de métodos complementares que oferecem suporte às políticas baseadas em evidências. Embora este livro enfoque nos métodos quantitativos de avaliação de impacto, começaremos colocando estes métodos no contexto mais amplo de resultados, o que também inclui monitoramento e outros tipos de avaliação.

O *monitoramento* é um processo contínuo, que acompanha o que está acontecendo com o programa e usa os dados coletados para informar a implementação do programa e a gestão e decisões tomadas no dia a dia. Utilizando, principalmente, dados administrativos, o monitoramento rastreia o desempenho do programa e o compara ao resultado esperado, realiza comparações entre programas e analisa tendências ao longo do tempo. Geralmente, o monitoramento cobre insumos, atividades e produtos, embora possa, às vezes, incluir resultados, tais como o progresso em direção às metas nacionais de desenvolvimento.

As *Avaliações* são processos periódicos, julgamentos objetivos de um projeto, programa ou política planejada, em andamento ou concluída.

## **Quadro 1.1: Avaliações e Sustentabilidade Política**

### **O programa de transferência condicionada de renda *Progres*a/Oportunidades no México**

Nos anos 90, o governo do México lançou um programa inovador de transferência condicionada de renda chamado “Progres”a”. Seu objetivo era dar às famílias pobres um apoio financeiro de curto prazo e criar incentivos para investimentos no capital humano das crianças, principalmente por meio da transferência de renda às mães de famílias pobres com a condição de que os seus filhos estejam frequentando regularmente a escola e comparecendo a um centro de saúde.

Desde o começo, o governo considerou que era fundamental monitorar e avaliar o programa. As autoridades do programa contrataram um grupo de pesquisadores para desenhar uma avaliação de impacto e integrá-la ao processo de expansão do programa, à medida que este fosse estendido às comunidades participantes.

As eleições presidenciais de 2000 trouxeram uma mudança de partido no poder. Em 2001, os avaliadores externos do “Progres”a” apresentaram os resultados de seu estudo ao governo recém-eleito. Os resultados do programa eram impressionantes: eles mostraram que o programa era bem focalizado nos pobres e que gerou mudanças promissoras no capital humano das famílias. Schultz (2004) concluiu que o programa aumentou

significativamente os anos de matrícula escolar, em uma média de 0,7 anos adicionais de estudo. Gertler (2004) constatou que a incidência de doenças infantis diminuiu 23%, enquanto os adultos relataram uma redução de 19% na quantidade de dias de doença ou incapacidade. Entre os resultados nutricionais, Behrman e Hoddinott (2001) relatam que o programa reduziu a probabilidade de baixo crescimento em cerca de 1 centímetro por ano para crianças na faixa etária crítica, de 12 a 36 meses.

Esses resultados da avaliação deram suporte a um diálogo político baseado em evidências e contribuíram para a decisão do novo governo de continuar com o programa. Por exemplo, o governo expandiu o alcance do programa, introduzindo bolsas para os alunos de nível médio e melhorando os programas de saúde para adolescentes. Ao mesmo tempo, os resultados foram usados para modificar outros programas assistenciais - como o subsídio *tortilla*, amplo e menos focalizado, que foi reduzido.

A avaliação bem-sucedida do Progres”a” contribuiu para a rápida adoção dos programas de transferência condicionada de renda em todo o mundo, assim como a adoção de legislação, no México, exigindo que todos os projetos sociais sejam avaliados.

*Fontes:* Behrman e Hoddinott 2001; Gertler 2004; Fiszbein e Schady 2009; Levy e Rodriguez 2005; Schultz 2004; Skoufias e McClafferty 2001.

## **Quadro 1.2: Avaliando para Melhorar a Alocação de Recursos**

### **Planejamento familiar e fertilidade na Indonésia**

Na década de 70, os esforços inovadores do planejamento familiar na Indonésia ganharam reconhecimento internacional por seu sucesso em diminuir as taxas de fertilidade do país. A notoriedade surgiu de dois fenômenos paralelos: (1) as taxas de fertilidade apresentaram uma queda, de 22% entre os anos de 1970 e 1980, de 25% entre os anos de 1981 e 1990 e um pouco mais moderada entre os anos de 1991 e 1994; e (2) durante o mesmo período, o governo da Indonésia aumentou consideravelmente os recursos alocados ao planejamento familiar (particularmente, os subsídios para anticoncepcionais). Dado que as duas coisas ocorreram simultaneamente, muitos concluíram que foi o aumento do investimento no planejamento familiar que levou a um índice de fertilidade mais baixo.

Não convencidos pela evidência disponível, uma equipe de pesquisadores testou se os programas de planejamento familiar realmente diminuíram as taxas de fertilidade. Eles descobriram, ao contrário do que se acreditava, que os programas de planejamento familiar tiveram um impacto apenas moderado sobre a fertilidade e argumentaram que, ao invés disso, foi a mudança no status das mulheres que teria sido responsável pelo declínio na taxa de fertilidade. Os pesquisadores notaram que, antes do início do programa de planejamento familiar, muito poucas mulheres na idade fértil haviam

concluído o ensino fundamental. Durante o mesmo período do programa de planejamento familiar, no entanto, o governo executou um amplo programa de educação para meninas, de modo que, ao final do programa, as mulheres que entravam na idade fértil haviam se beneficiado de mais anos de educação. Quando o crescimento do setor petrolífero expandiu a economia e aumentou a demanda por trabalho na Indonésia, a participação de mulheres escolarizadas no mercado de trabalho cresceu consideravelmente. Conforme aumentou o valor do tempo de trabalho da mulher, aumentou também o uso de anticoncepcionais. Ao final, salários mais altos e empoderamento das mulheres foram responsáveis por 70% do declínio observado na fertilidade, mais do que o investimento em programas de planejamento familiar.

Estes resultados de avaliação informaram as decisões sobre a alocação de recursos tomadas posteriormente pelos gestores de políticas: o financiamento foi reprogramado, dos subsídios aos anticoncepcionais para os programas de estímulo à matrícula escolar das mulheres. Embora os objetivos finais dos dois tipos de programa fossem semelhantes, os estudos de avaliação demonstraram que, no contexto da Indonésia, taxas mais baixas de fertilidade poderiam ser alcançadas de modo mais eficiente através de investimentos em educação, ao invés de investimentos em planejamento familiar.

*Fontes:* Gertler e Molyneaux 1994, 2000.

As avaliações são usadas para responder a perguntas específicas relacionadas à formulação, implementação e obtenção de resultados. Ao contrário do monitoramento contínuo, elas são realizadas em pontos discretos no tempo e, geralmente, buscam uma perspectiva externa de especialistas. O seu desenho, método e custo variam substancialmente, dependendo do tipo de pergunta que a avaliação tenta responder. Em geral, as avaliações podem abordar três tipos de questões (Imas e Rist 2009):

- *Perguntas descritivas.* A avaliação procura determinar o que está sendo executado e descreve processos, condições, relações organizacionais e pontos de vista das partes interessadas.
- *Perguntas normativas.* A avaliação compara o que está sendo executado ao que deveria ocorrer; avalia atividades e se as metas estão ou não sendo alcançadas. As perguntas normativas podem ser aplicadas a insumos, atividades ou produtos.
- *Perguntas de causa e efeito.* A avaliação analisa resultados e procura identificar que diferença a intervenção faz para os resultados.

*As avaliações de impacto* são um tipo particular de avaliação, que procura responder a perguntas de causa e efeito. Diferentemente das avaliações gerais, que podem responder a muitos tipos de perguntas, as avaliações de impacto se estruturam em torno de um tipo específico de pergunta: *qual é o impacto (ou efeito causal) de um programa sobre um resultado de interesse?* Esta pergunta básica incorpora uma importante dimensão causal: estamos interessados apenas no *impacto* do programa, isto é, no efeito que o programa causa diretamente no resultado. Uma avaliação de impacto busca as mudanças nos resultados que são *atribuíveis diretamente ao programa*.

O enfoque na causalidade e atribuição de resultados é a característica distintiva das avaliações de impacto e determina as metodologias que podem ser usadas. Para poder estimar o efeito causal ou impacto de um programa sobre os resultados, qualquer método escolhido deve estimar o chamado *contrafactual*, isto é, qual teria sido o resultado para os participantes do programa se eles não tivessem participado do programa. Na prática, a avaliação de impacto exige que o avaliador encontre um grupo de comparação para estimar o que teria acontecido aos participantes do programa na ausência do programa. A segunda parte do livro descreve os principais métodos que podem ser usados para encontrar grupos de comparação adequados.

A pergunta básica da avaliação de impacto, *“qual é o impacto (ou efeito causal) de um programa sobre um resultado de interesse?”* pode ser aplicada a muitos contextos. Por exemplo, qual é o efeito causal das bolsas na frequência escolar e desempenho acadêmico dos alunos? Qual é o impacto no acesso

**Conceito-chave:**

A pergunta básica da avaliação de impacto pode ser formulada como: qual é o impacto (ou efeito causal) de um programa sobre um resultado de interesse?

aos serviços de saúde da contratação de serviços de atenção básica de fornecedores privados? Se o chão de terra batida fosse substituído por chão de cimento, qual seria o impacto na saúde das crianças? Será que estradas melhores aumentam o acesso aos mercados de trabalho e a renda familiar? Caso aumentem, em quanto? O tamanho das turmas influencia o desempenho dos alunos? Caso influencie, em quanto? O que é mais efetivo para o aumento do uso de mosquiteiros em áreas afetadas pela malária, campanhas pelo correio ou sessões de treinamento?

## **Avaliação de Impacto em Decisões de Políticas**

As avaliações de impacto são necessárias para informar os formuladores de políticas em várias decisões, desde a redução de programas ineficientes até a expansão de intervenções que funcionam, a adequação de benefícios do programa e a seleção entre várias alternativas de programas. Elas são mais efetivas quando aplicadas seletivamente para responder a perguntas importantes de políticas, e podem ser particularmente efetivas quando aplicadas a programas piloto inovadores que avaliam uma nova abordagem ainda não testada, porém promissora. A avaliação do programa mexicano Progreso/Oportunidades, descrita no Quadro 1.1, se tornou bastante influente, não somente devido à natureza inovadora do programa, mas também porque sua avaliação de impacto trouxe evidências fortes e convincentes, que não poderiam ser ignoradas em decisões de política subsequentes. A adoção e a extensão do programa foram fortemente influenciadas pelos resultados da avaliação. Hoje, o programa Oportunidades atinge cerca de 1 em cada 4 mexicanos e é uma peça central da estratégia do México para combater a pobreza.

As avaliações de impacto podem ser usadas para explorar diferentes tipos de questões sobre políticas. A forma básica de avaliação de impacto testará a efetividade de um determinado programa. Em outras palavras, responderá à pergunta: *determinado programa é efetivo quando comparado à sua própria ausência?* Conforme apresentado na segunda parte, este tipo de avaliação de impacto depende da comparação de um grupo de tratamento que recebeu o projeto, programa ou política com um grupo de comparação que não o recebeu, a fim de estimar a efetividade do programa.

Além de responder à pergunta básica da avaliação, as avaliações também podem ser usadas para testar a efetividade das alternativas de implementação do programa - isto é, para responder à pergunta: *quando um programa puder ser implementado de várias formas, qual delas será a mais efetiva?* Neste tipo de avaliação, duas ou mais abordagens dentro de um mesmo programa podem ser comparadas para gerar evidências sobre qual é a melhor

alternativa para se atingir um objetivo específico. Estas alternativas de programa são, geralmente, chamadas de “braços de tratamento”. Por exemplo, quando a quantidade de benefícios que um programa deve fornecer não estiver clara (20 ou 80 horas de capacitação?), as avaliações de impacto podem testar o impacto relativo de intensidades variadas de tratamento (veja um exemplo no Quadro 1.3). Avaliações de impacto que testam tratamentos alternativos de programas geralmente incluem um grupo de tratamento para cada um dos tipos (ou “braços”) de tratamento, bem como um

### **Quadro 1.3: Avaliando para Melhorar o Desenho do Programa Desnutrição e desenvolvimento cognitivo na Colômbia**

No início dos anos 70, a Estação de Pesquisa de Ecologia Humana implementou, em colaboração com o Ministério da Educação colombiano, um programa piloto para enfrentar a situação da desnutrição infantil em Cali, na Colômbia, provendo cuidados à saúde e atividades educativas, assim como suplementos nutricionais e alimentícios. Como parte do piloto, uma equipe de avaliadores foi encarregada de determinar (1) quanto tempo o programa deveria durar para reduzir a desnutrição entre crianças de idade pré-escolar em famílias de baixa renda e (2) se as intervenções também poderiam resultar em melhorias no desenvolvimento cognitivo.

O programa foi disponibilizado para todas as famílias elegíveis e, durante o piloto, os avaliadores puderam comparar grupos semelhantes de crianças que receberam diferentes durações do tratamento. Os avaliadores usaram, primeiramente, um processo de filtragem, para identificar um grupo-alvo de 333 crianças desnutridas. Estas crianças foram, então, classificadas por vizinhança em 20 setores; cada setor foi aleatoriamente designado a um dos quatro grupos de tratamento. Os grupos se diferenciavam somente

na sequência em que começavam o tratamento e, portanto, na quantidade de tempo que permaneceram no programa. O grupo 4 começou antes e foi exposto ao tratamento pelo período mais longo, seguido pelos grupos 3, 2 e, por último, 1. O tratamento em si consistiu em 6 horas diárias de cuidados com a saúde e atividades educativas, além de alimentos adicionais e suplementos nutricionais. Em intervalos regulares durante o programa, os avaliadores usaram testes cognitivos para acompanhar o progresso das crianças em todos os quatro grupos.

Os avaliadores descobriram que as crianças que permaneceram no programa por mais tempo demonstraram os maiores ganhos em melhoria cognitiva. No teste de inteligência de Stanford-

Binet, que estima a idade mental menos a idade cronológica, o grupo 4 obteve uma média de -5 meses e o grupo 1 obteve uma média de -15 meses.

Este exemplo ilustra como os implementadores de programas e formuladores de políticas podem usar as avaliações de múltiplos tratamentos para determinar a alternativa mais efetiva de programa.

*Fonte: McKay et al. 1978.*



grupo de comparação “limpo” e que não recebe nenhuma intervenção do programa. As avaliações de impacto também podem ser usadas para testar inovações ou alternativas de implementação de um determinado programa. Por exemplo, um programa pode desejar testar alternativas de campanhas de sensibilização - selecionando um grupo para receber a campanha pelo correio enquanto os outros recebem visitas de casa em casa - para avaliar qual é a mais efetiva.

## **Decidindo se Deve ou não Avaliar**

Nem todos os programas justificam uma avaliação de impacto. As avaliações de impacto podem ser onerosas e o orçamento para tal deve ser usado estrategicamente. Se você estiver iniciando ou considerando expandir um novo programa e estiver refletindo se é o momento de ir em frente com uma avaliação de impacto, algumas perguntas básicas ajudarão na decisão.

A primeira pergunta a ser feita é: *O que está em questão no programa?* A resposta a esta pergunta dependerá tanto do orçamento como da quantidade de pessoas que são ou serão afetadas pelo programa. Consequentemente, as próximas questões são: *o programa requer ou requererá uma grande parte do orçamento disponível? O programa afeta ou afetará um grande número de pessoas?* Se o programa não requer orçamento ou afetar somente algumas pessoas, pode ser que não valha a pena avaliar. Por exemplo, no caso de um programa que fornece aconselhamento a pacientes em hospitais por meio de voluntários, o orçamento envolvido e a quantidade de pessoas afetadas pode não justificar uma avaliação de impacto. Por outro lado, uma reforma salarial para o magistério, que eventualmente venha a impactar todos os professores do Ensino Fundamental de um determinado país, caracteriza um programa com riscos muito mais altos.

Se você determinar que o programa é suficientemente significativo então, a próxima pergunta é se há alguma evidência que mostre que o programa funciona. Em particular, você sabe qual seria o tamanho do impacto do programa? Há evidências disponíveis de um país semelhante com circunstâncias semelhantes? Se não houver nenhuma evidência disponível a respeito do potencial do programa sob consideração, você pode querer iniciar um piloto que incorpore uma avaliação de impacto. Por outro lado, se existirem evidências disponíveis envolvendo circunstâncias semelhantes, o custo de uma avaliação de impacto provavelmente só se justificará se puder abordar uma questão de política nova e relevante. Este seria o caso se o seu programa incluísse inovações importantes e que ainda não puderam ser testadas.

Para justificar a mobilização dos recursos técnicos e financeiros necessários para executar uma avaliação de impacto de alta qualidade, o programa a ser avaliado deve ser:

- *Inovador*: está testando uma abordagem nova e promissora.
- *Replicável*: o programa pode ser expandido ou pode ser aplicado em um contexto diferente.
- *Estrategicamente relevante*: o programa é uma iniciativa emblemática; requer recursos consideráveis; cobre - ou poderia ser expandido para cobrir - um grande número de pessoas; ou poderia gerar grandes economias.
- *Ainda não testado*: pouco se sabe sobre a efetividade do programa, em contexto global ou particular.
- *Influente*: os resultados serão usados para informar decisões cruciais sobre políticas.

## Análise de Custo-Efetividade

Uma vez que os resultados da avaliação de impacto estejam disponíveis, eles podem ser combinados com informações sobre os custos do programa, para responder a duas perguntas adicionais. Primeiramente, considerando a forma básica da avaliação de impacto, adicionar informações de custo nos permitirá realizar uma análise de custo-benefício, que responderá à pergunta: *qual é a relação custo-benefício de um determinado programa?* A análise custo-benefício estima o total de benefícios esperados de um programa, comparado ao total de custos esperados. Ela procura quantificar, em termos monetários, todos os custos e benefícios de um programa e avaliar se os benefícios ultrapassaram os custos.

Em um mundo ideal, a análise de custo-benefício baseada em evidências de avaliação de impacto ocorreria não somente para um programa específico, mas também para uma série de programas ou alternativas de programa, de modo que os formuladores de políticas pudessem avaliar qual programa ou alternativa de programa teria a melhor relação custo-efetividade para alcançar um determinado objetivo. Quando uma avaliação de impacto testa alternativas de programa, a adição de informações de custo nos permite responder uma segunda pergunta: *como as várias alternativas de implementação do programa se comparam, em termos de custo-efetividade?* Esta análise de custo-efetividade compara o desempenho relativo de dois ou mais programas (ou alternativas de programa) para alcançar um mesmo resultado.

### Conceito-chave:

A análise custo-benefício estima o total de benefícios esperados de um programa, comparado ao total de custos esperados.

### Conceito-chave:

A análise de custo-efetividade compara o desempenho relativo de dois ou mais programas - ou de alternativas de um programa para alcançar um mesmo resultado.

Em uma análise de custo-benefício ou custo-efetividade, a avaliação de impacto estima o lado do benefício e da efetividade e a análise de custos fornece informações sobre o custo. Este livro enfoca a avaliação de impacto e não discute detalhadamente como coletar dados de custos ou realizar a análise de custo-benefício<sup>2</sup>. Entretanto, é fundamental que a avaliação de impacto seja complementada por informações sobre os custos do projeto, do programa ou da política que estiver sendo avaliada. Uma vez que as informações de impactos e custos estejam disponíveis para uma variedade de programas, a análise de custo-efetividade pode identificar quais investimentos geram as taxas mais altas de retorno e permite aos gestores de políticas tomarem decisões informadas sobre as intervenções em que devem investir. O Quadro 1.4 ilustra como as avaliações de impacto podem ser usadas para identificar os programas de melhor custo-efetividade e melhorar a alocação de recursos.

#### **Quadro 1.4: Avaliando a Custo-Efetividade** **Comparando estratégias para aumentar a frequência escolar no Quênia**

Ao avaliar uma variedade de programas em um contexto semelhante, é possível comparar o custo-efetividade relativo de alternativas que buscam melhorar resultados tais como a frequência escolar. No Quênia, a organização não governamental *International Child Support Africa* (ICS Africa) implementou uma série de intervenções educacionais que incluíam tratamento contra parasitas intestinais e uniformes e merenda escolar gratuitos. Cada uma das intervenções passou por uma avaliação de natureza aleatória e uma análise de custo benefício - a comparação entre elas traz conhecimentos interessantes sobre como aumentar a frequência escolar.

O programa de medicamentos contra parasitas intestinais para crianças em idade escolar aumentou a frequência em aproximadamente 0,14 anos por criança tratada, a um custo estimado de \$0,49 por criança. Isto equivale a aproximadamente \$3,50 por ano

adicional de participação escolar, incluindo as externalidades experimentadas por crianças e adultos que, embora não estejam na escola, vivem nas comunidades beneficiadas pela redução na transmissão de parasitas.

Uma segunda intervenção, o programa de patrocínio infantil, reduziu o custo da frequência escolar ao fornecer uniformes escolares a alunos em sete escolas selecionadas aleatoriamente. As taxas de evasão caíram dramaticamente nas escolas tratadas e, após 5 anos, estima-se que o programa aumente em 17%, em média, os anos de matrícula escolar. No entanto, mesmo sob as hipóteses mais otimistas, o custo de aumentar a frequência escolar usando o programa de uniforme escolar foi estimado em aproximadamente \$99 por ano adicional de frequência escolar.

Finalmente, o programa de café da manhã gratuito oferecido a crianças em 25 pré-escolas

*(continua)*

### Quadro 1.3 *continuação*

selecionadas aleatoriamente levou a um aumento de 30% na frequência nas escolas de tratamento, a um custo estimado de \$36 por ano adicional de escolarização. Os resultados dos testes de aprendizagem também aumentaram em 0,4 desvios-padrão, contanto que os professores tenham sido bem treinados antes do programa.

Embora intervenções semelhantes possam ter objetivos diferentes - tais como os efeitos do tratamento das verminoses na saúde ou desempenho escolar devido a uma maior frequência escolar - comparar uma série de avaliações realizadas no mesmo contexto pode revelar quais programas alcançaram o objetivo desejado, com os menores custos.

*Fontes:* Kremer e Miguel, 2004; Kremer, Moulin e Namunyu, 2003; Poverty Action Lab, 2005; Vermeersch e Kremer, 2005.

## Avaliação Prospectiva versus Avaliação Retrospectiva

### **Conceito-chave:**

Avaliações prospectivas são desenvolvidas ao mesmo tempo em que o programa está sendo elaborado e são integradas à implementação do programa

As avaliações de impacto podem ser divididas em duas categorias: prospectivas e retrospectivas. Avaliações prospectivas são desenvolvidas ao mesmo tempo em que o programa está sendo elaborado e são integradas à implementação do programa. Dados de linha de base são coletados antes da implementação do programa, tanto para grupos de tratamento quando de comparação. As avaliações retrospectivas avaliam o impacto do programa após sua implementação, gerando grupos de tratamento e de comparação *a posteriori*.

Em geral, as avaliações de impacto prospectivas têm maior probabilidade de produzir resultados mais robustos e confiáveis, por três razões.

Em primeiro lugar, podem ser coletados dados de linha de base para estabelecer as medidas pré-programa dos resultados de interesse. Os dados de linha de base fornecem informações sobre beneficiários e grupos de comparação antes da implementação do programa e são importantes para medir os resultados pré-intervenção. Devem ser analisados os dados de linha de base dos grupos de tratamento e de comparação para garantir que os grupos sejam semelhantes. As linhas de base também podem ser usadas para avaliar a efetividade da focalização - isto é, se o programa vai, ou não, atingir os beneficiários pretendidos.

Em segundo lugar, definir medidas de sucesso de um programa na etapa de planejamento ajuda a focar a avaliação e o próprio programa nos resultados pretendidos. Como veremos, as avaliações de impacto têm origem na teoria de mudança do programa ou cadeia de resultados. O desenho de uma avaliação de impacto ajuda a esclarecer os objetivos do programa,

em particular porque requer o estabelecimento de medidas bem definidas do sucesso do programa. Os formuladores de políticas devem estabelecer objetivos claros e perguntas de avaliação para garantir que os resultados sejam altamente relevantes em termos de política. Na verdade, o apoio integral dos formuladores de políticas é um pré-requisito para executar uma avaliação bem-sucedida; as avaliações de impacto não devem ser realizadas a menos que os formuladores de políticas estejam convencidos da legitimidade da avaliação e de seu valor, no sentido de informar decisões sobre políticas.

Em terceiro lugar, o mais importante: em uma avaliação prospectiva, os grupos de tratamento e de comparação são identificados antes que o programa seja implementado. Como explicaremos mais detalhadamente nos capítulos seguintes, existem muito mais opções para executar avaliações válidas quando as avaliações são planejadas desde o princípio e integradas à implementação do projeto. Nas partes 2 e 3, discutimos que quase sempre pode ser encontrada uma estimativa válida do cenário contrafactual para qualquer programa que tenha regras de atribuição claras e transparentes, desde que a avaliação seja elaborada prospectivamente. Em suma, as avaliações prospectivas têm maiores chances de gerar contrafatuais válidos. Na fase de desenho, podem ser consideradas formas alternativas de estimar-se um contrafactual válido. O desenho da avaliação de impacto também pode ser integralmente alinhado às regras de operação do programa, assim como às etapas de implementação e expansão do programa.

Por outro lado, nas avaliações retrospectivas, o avaliador geralmente possui informações tão limitadas que se torna difícil analisar se o programa foi implementado com sucesso ou se os seus participantes realmente extraíram benefícios dele. Em parte, a razão disto se deve ao fato de que muitos programas não coletam dados de linha de base a menos que a avaliação tenha sido integrada desde o início e, uma vez que o programa começa a funcionar, é tarde demais para isso.

As avaliações retrospectivas que usam dados existentes são necessárias para avaliar os programas que foram alocados no passado. De modo geral, as opções para se obter uma estimativa válida do cenário contrafactual são muito mais limitadas nessas situações. A avaliação depende de regras claras de operação do programa referentes à alocação dos benefícios. Também depende da disponibilidade de dados, com cobertura suficiente dos grupos de tratamento e de comparação, tanto antes como depois da implementação do programa. Como resultado, a possibilidade de uma avaliação retrospectiva depende do contexto e nunca é garantida. Mesmo quando são possíveis de realizar, as avaliações retrospectivas geralmente lançam mão de métodos quase-experimentais, que dependem de fortes pressupostos; assim, elas tendem a produzir evidências mais facilmente contestáveis.

## Estudos de Eficácia e de Efetividade

O principal papel da avaliação de impacto é produzir evidências sobre a efetividade de um programa para o uso de autoridades governamentais, gestores de programas, sociedade civil e outras partes interessadas. Os resultados da avaliação de impacto são particularmente úteis quando as conclusões podem ser aplicadas de forma mais ampla à população de interesse. A questão de generalização (conhecida como “validade externa” na literatura de métodos de pesquisa) é essencial para os tomadores de decisão, pois determina se os resultados identificados na avaliação podem ser replicados para outros grupos além daqueles estudados na avaliação, caso o programa venha a ser ampliado.

Nos primórdios das avaliações de impacto de programas de desenvolvimento, uma grande parcela de evidências se baseava em *estudos de eficácia* realizados sob circunstâncias muito específicas; infelizmente, os resultados daqueles estudos não podiam, com frequência, ser generalizados para além do escopo da avaliação. Os estudos de eficácia são, geralmente, realizados em um cenário muito específico, com forte envolvimento técnico de pesquisadores durante a execução do programa. Estes estudos de eficácia geralmente são realizados para provar conceitos e testar a viabilidade de um novo programa. Se o programa não gera os impactos antecipados sob tais condições, geralmente administradas com muito cuidado, é improvável que funcione sob circunstâncias normais. Por serem geralmente executados sob condições rigorosamente administradas, os impactos de tais pilotos de pequena escala podem não ser necessariamente informativos sobre o impacto de um projeto similar executado em uma escala maior e sob circunstâncias normais. Por exemplo, uma intervenção piloto que introduza novos protocolos de tratamento pode funcionar em um hospital com administradores e equipe médica excelentes, mas a mesma intervenção pode não funcionar em um hospital mediano, com administradores menos atentos e equipe limitada. Além disso, os cálculos de custo-benefício irão variar, considerando que os custos fixos e as economias de escala podem não ser captados em estudos de eficácia de pequena escala. Consequentemente, embora as evidências dos estudos de eficácia possam ser úteis para testar novas abordagens, os resultados geralmente apresentam validade externa limitada e nem sempre representam adequadamente os contextos mais gerais que, normalmente, configuram a principal preocupação dos gestores de políticas.

Por outro lado, os *estudos de efetividade* fornecem evidências a partir de intervenções que ocorrem em circunstâncias normais, pelo uso de canais regulares de execução. Quando as avaliações de efetividade são desenhadas e executadas adequadamente, os resultados obtidos se confirmarão não

somente para a amostra da avaliação, mas também para outros potenciais beneficiários fora da amostra. A validade externa é de fundamental importância para os formuladores de políticas, porque lhes permite usar os resultados da avaliação para informar decisões sobre o programa aplicáveis aos beneficiários além da amostra da avaliação.

## **Combinando Fontes de Informação para Avaliar “o Quê” e “Por Quê”**

As avaliações de impacto realizadas isoladamente de outras fontes de informações são vulneráveis, tanto tecnicamente quanto em termos da sua efetividade potencial. Sem informações sobre a natureza e o conteúdo do programa para contextualizar os resultados da avaliação, os gestores de políticas ficam em dúvida em relação ao porquê de determinados resultados terem, ou não, sido atingidos. Enquanto as avaliações de impacto produzem estimativas confiáveis dos efeitos causais de um programa, elas não são elaboradas especificamente para gerar conhecimentos sobre a implementação do programa. Além disso, as avaliações de impacto devem ser bem alinhadas à implementação do programa e, conseqüentemente, precisam ser orientadas por informações sobre como, quando e onde o programa avaliado está sendo implementado.

Dados qualitativos, de monitoramento e avaliações de processos são necessários para rastrear a implementação do programa e examinar questões de processo que sejam críticas para informar e interpretar os resultados das avaliações de impacto. Neste sentido, as avaliações de impacto e outras formas de avaliação são complementares entre si, não substitutas.

Por exemplo, um governo provincial pode decidir anunciar que pagará bônus às clínicas médicas da zona rural se elas aumentarem a porcentagem de partos assistidos por um profissional da saúde na clínica. Se a avaliação concluir que não ocorreu mudança alguma na porcentagem de partos assistidos na clínica, podem existir muitas explicações possíveis e a cada uma corresponde uma ação necessária. Primeiro, pode ser que as equipes das clínicas rurais não possuam informações suficientes sobre os bônus ou não entendam as regras do programa. Neste caso, o governo provincial precisaria reforçar as informações sobre o bônus, além de oferecer campanhas educativas aos centros de saúde. Alternativamente, se a falta de equipamentos ou corte de energia impedir que as clínicas médicas recebam mais pacientes, pode ser necessário melhorar o sistema de suporte e melhorar o fornecimento de energia. Finalmente, as gestantes da zona rural podem não querer usar as clínicas; pode ser que elas prefiram parteiras e o parto feito em casa, por

razões culturais. Neste caso, pode ser mais eficaz derrubar as barreiras que as mulheres enfrentam em acessar as clínicas do que dar bônus às clínicas. Portanto, uma boa avaliação de impacto permitirá ao governo determinar se a taxa de partos assistidos mudou como resultado do programa de bônus. No entanto, são necessárias abordagens de avaliação complementares para entender se o programa foi implementado conforme o planejamento e onde faltam conexões. Neste exemplo, os avaliadores desejariam complementar sua análise de impacto entrevistando o pessoal da clínica de saúde em relação a seu conhecimento do programa, revisar a disponibilidade de equipamento nas clínicas, realizar grupos focais com gestantes para entender suas preferências e barreiras ao acessar o serviço clínico e examinar todos os dados disponíveis sobre o acesso aos centros de saúde na zona rural.

## **Usando Dados Qualitativos**

*Dados qualitativos* são um suplemento-chave para as avaliações de impacto quantitativas, pois podem fornecer perspectivas complementares sobre o desempenho do programa. As avaliações que integram análises qualitativas e quantitativas são caracterizadas pelo uso de “métodos combinados” (Bamberger, Raim e Woolcock 2010). Abordagens qualitativas incluem grupos focais e entrevistas com beneficiários selecionados ou outros informantes-chave (Rai e Woolcock 2003). Embora os pontos de vista e opiniões reunidos durante as entrevistas e grupos focais possam não ser representativos dos beneficiários do programa, eles são particularmente úteis durante as três etapas de uma avaliação de impacto:

1. Durante o desenho da avaliação de impacto, os avaliadores podem usar grupos focais e entrevistas com informantes-chave para desenvolver hipóteses sobre como e por que o programa funcionaria e esclarecer questões de pesquisa que precisem ser abordadas no trabalho de avaliação quantitativa de impacto.
2. Na etapa intermediária, antes que os resultados da avaliação quantitativa de impacto estejam disponíveis, o trabalho qualitativo pode ajudar a fornecer aos gestores de políticas panoramas rápidos do que vem acontecendo no programa.
3. Na etapa de análise, os avaliadores podem aplicar métodos qualitativos para contextualizar e explicar os resultados quantitativos, para explorar casos “aberrantes” de sucesso e fracasso e desenvolver explicações sistemáticas do desempenho do programa, conforme expresso pelos resultados quantitativos. Neste sentido, o trabalho qualitativo pode



ajudar a explicar por que são observados determinados resultados na análise quantitativa e pode ser usado para examinar a “caixa preta” do que acontece no programa (Bamberger, Rao e Woolcock 2010).

## Usando Dados de Monitoramento e Avaliações de Processo

Os dados de monitoramento também são um recurso crítico em uma avaliação de impacto. Eles permitem ao avaliador verificar quais participantes receberam o programa, com que rapidez o programa está se expandindo, como os recursos estão sendo gastos e, de maneira geral, se as atividades estão sendo implementadas conforme planejado. Estas informações são fundamentais ao implementar a avaliação, por exemplo, para garantir que os dados de linha de base sejam coletados antes que o programa seja introduzido e para verificar a integridade dos grupos de tratamento e de comparação. Além disso, o sistema de monitoramento pode fornecer informações sobre o custo de implementação do programa, que também são necessárias para a análise de custo-benefício.

Finalmente, *as avaliações de processo* enfocam como um programa é implementado e operado, avaliando se ele está em conformidade com o desenho original, e documenta seu desenvolvimento e operação. As avaliações de processos podem, geralmente, ser efetuadas rapidamente e a um custo razoável. Nos pilotos e nas etapas iniciais de um programa, elas podem ser uma fonte valiosa de informação sobre como melhorar a implementação do programa.

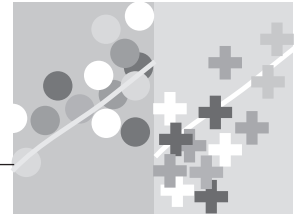
## Notas

1. Vide Fiszbein e Schady (2009) para uma visão geral dos programas de CCT e o papel influente desempenhado pelo Progres/Oportunidades devido à sua avaliação de impacto.
2. Para uma discussão detalhada da análise custo-benefício, vide Belli et al. 2001; Boardman et al. 2001; Brent 1996; ou Zerbe e Dively 1994.

## Referências

- Bamberger, Michael, Rao & Woolcock. (2010). Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development. *Policy Research Working Paper 5245*. Washington, DC: Banco Mundial.
- Behrman, J. & Hoddinott, J. (2001). An Evaluation of the Impact of PROGRESA on Pre-school Child Height. *FCND Briefs 104*. Washington, DC: Instituto Internacional de Pesquisa Política em Alimentos.

- Belli, P., Anderson, J., Barnum, H., Dixon, J. & Tan, J. (2001). *Handbook of Economic Analysis of Investment Operations*. Washington, DC: Banco Mundial.
- Boardman, A., Vining, A., Greenberg, D. & Weimer, D. (2001). *Cost-Benefit Analysis: Concepts and Practice*. New Jersey: Prentice Hall.
- Brent, Robert. 1996. *Applied Cost-Benefit Analysis*. England: Edward Elgar.
- Fiszbein, A. & Schady, N. (2009). *Conditional Cash Transfer, Reducing Present and Future Poverty*. Relatório de Pesquisa Política do Banco Mundial. Washington, DC: Banco Mundial.
- Gertler, Paul J. 2004. Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment. *American Economic Review* 94 (2): 336–41.
- Gertler, Paul J. & Molyneaux, J.M. (1994). How Economic Development and Family Planning Programs Combined to Reduce Indonesian Fertility. *Demography* 31 (1): 33–63.
- . (2000). The Impact of Targeted Family Planning Programs in Indonesia. *Population and Development Review* 26: 61–85.
- Imas, L. & Rist, R. (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: Banco Mundial.
- Kremer, M. & Miguel, E. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica* 72 (1): 159–217.
- Kremer, M., Moulin, S. & Namunyu, R. (2003). Decentralization: A Cautionary Tale. Poverty Action Lab Paper 10, Cambridge, MA: Instituto de Tecnologia de Massachusetts.
- Levy, S. & Rodríguez, E. (2005). *Sin Herencia de Pobreza: El Programa Progresas-Oportunidades de México*. Washington, DC: Banco de Desenvolvimento Interamericano.
- McKay, H., McKay, A., Siniestra, L., Gomez, H. & Lloreda, P. (1978). Improving Cognitive Ability in Chronically Deprived Children. *Science* 200 (21): 270–78.
- Poverty Action Lab. 2005. Primary Education for All. *Fighting Poverty: What Works?* 1 (Fall): n.p., disponível em: <http://www.povertyactionlab.org>.
- Rao, V. & Woolcock, M. (2003). Integrating Qualitative and Quantitative Approaches in Program Evaluation. In *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*, ed. Bourguignon, F.J. & Pereira da Silva, L. 165–90. Nova York: Oxford University Press.
- Schultz, P. (2004). School Subsidies for the Poor: Evaluating the Mexican Progresas Poverty Program. *Journal of Development Economics* 74 (1): 199–250.
- Skoufias, E. & McClafferty, B. (2001). *Is Progresas Working? Summary of the Results of an Evaluation by IFPRI*. Washington, DC: International Food Policy Research Institute.
- Vermeersch, C. & Kremer, M. (2005). School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation. *Documentos de Trabalho de Relatório Político* 3523. Washington, DC: Banco Mundial.
- Zerbe, R., & Dively, D. (1994). *Benefit Cost Analysis in Theory and Practice*. Nova York: Harper Collins Publishing.



---

## CAPÍTULO 2

# Determinando as Perguntas de Avaliação

Este capítulo define os passos iniciais na elaboração de uma avaliação. Os passos incluem a definição do tipo de pergunta a ser respondida pela avaliação, a construção da teoria de mudança, que detalha como o projeto deve alcançar os resultados pretendidos, o desenvolvimento de uma cadeia de resultados, a formulação das hipóteses a serem testadas pela avaliação e a seleção dos indicadores de desempenho.

Todos estes passos contribuem para determinar a pergunta de avaliação e devem ser dados no início do programa, envolvendo uma gama de atores relevantes - dos formuladores de políticas aos administradores do programa - com o objetivo de estabelecer uma visão comum dos objetivos do programa e de como eles serão atingidos. Tal envolvimento gera consensos em relação às principais questões a serem respondidas e fortalece os laços entre a avaliação, a execução do programa e a política. A aplicação dos passos fornece clareza e especificidade, úteis tanto para o desenvolvimento de uma boa avaliação de impacto como para o desenho e implementação de um programa efetivo. Cada passo - desde a clara especificação de objetivos e perguntas, passando pela articulação de ideias embutidas na teoria de mudança, até os resultados que o programa espera gerar - é claramente definido e articulado dentro do modelo lógico expresso na cadeia de resultados.

## Tipos de Perguntas de Avaliação

Toda avaliação começa com a formulação de uma pergunta de estudo, que dá foco à pesquisa e que é ajustada de acordo com o interesse da política em questão. Assim, avaliação consiste em gerar evidências confiáveis para que essa pergunta seja respondida. Como explicaremos abaixo, a pergunta básica de uma avaliação de impacto pode ser formulada tal como: *qual é o impacto ou efeito causal do programa sobre os resultados de interesse?* Em um exemplo que aplicaremos em toda a Parte 2, a questão de estudo é: *qual é o efeito do programa de Subsídio ao Seguro Saúde sobre os gastos das famílias com saúde?* A pergunta também pode ser orientada para testar opções de programa: *que combinação de campanhas pelo correio e orientação familiar funciona melhor para estimular o aleitamento materno como única fonte de alimentação?* Uma questão clara de avaliação é o ponto de partida de qualquer avaliação efetiva.

## Teorias de Mudança

Uma teoria de mudança é a descrição de como uma intervenção é pensada para gerar os resultados desejados. Ela descreve a lógica causal de como e por que um projeto, programa ou política alcançará os resultados pretendidos. A teoria de mudança é a chave subjacente de qualquer avaliação de impacto, dado o enfoque de causa e efeito da pesquisa. Como um dos primeiros passos de um desenho de avaliação, uma teoria de mudança pode ajudar a especificar as questões da pesquisa.

As teorias de mudança descrevem uma sequência de eventos que levam aos resultados; elas exploram as condições e pressupostos necessários para que a mudança ocorra, tornam explícita a lógica causal por trás do programa e mapeiam as intervenções do programa ao longo de trajetórias lógico-causais. Trabalhar com os atores relevantes do programa para construir uma teoria de mudança pode esclarecer e melhorar o desenho do programa. Isso é especialmente importante em programas que procuram influenciar o comportamento: teorias de mudança podem ajudar a desvincular os insumos e as atividades que compõem as intervenções do programa, os produtos a serem entregues e os resultados que derivam das mudanças comportamentais esperadas entre os beneficiários.

O melhor momento para desenvolver uma teoria de mudança para o programa é no início do processo de desenho, quando as partes interessadas podem reunir-se para desenvolver uma visão comum para o programa, seus objetivos e os caminhos para que esses objetivos sejam atingidos. As partes

interessadas podem, então, começar a execução do programa a partir de um entendimento comum do mesmo, de seu funcionamento e dos objetivos.

Além disso, os formuladores do programa devem revisar a literatura para buscar relatos de experiências sobre programas semelhantes e devem verificar os contextos e pressupostos por trás dos caminhos causais na teoria de mudança que eles estabelecem. Por exemplo, no caso do projeto de chão de concreto, no México, descrito no Quadro 2.1, a literatura fornecerá

### **Quadro 2.1: Teoria de Mudança**

#### **Do chão de cimento à felicidade no México**

Na sua avaliação do projeto “Piso Firme” (em espanhol) ou “Chão Firme”, Cattaneo et al. (2009) examinou o impacto da melhoria da moradia na saúde e bem-estar. Tanto o projeto quanto a avaliação foram motivados por uma teoria clara de mudança.

O objetivo do projeto “Piso Firme” é melhorar os padrões de vida, especialmente a saúde dos grupos vulneráveis vivendo em regiões de baixa renda e densamente povoadas do México. O projeto começou no estado de Coahuila, localizado ao norte do país, e se baseava em uma avaliação situacional realizada pelo governador Enrique Martínez e sua equipe de campanha.

A cadeia de resultados do programa é clara. Foi feita uma sondagem de porta em porta nos bairros elegíveis e foram oferecidos às famílias até 50 metros quadrados de cimento. O governo compra e entrega o cimento e as famílias e os voluntários da comunidade contribuem com a mão-de-obra para instalar o chão. O produto é a construção do chão de cimento, que pode ser concluída em cerca de um dia. Os resultados esperados da melhoria do ambiente doméstico incluem limpeza, saúde e felicidade.

A lógica desta cadeia de resultados é que os chãos de terra batida são um vetor de

parasitas, porque este tipo de chão dificilmente se mantém limpo. Os parasitas vivem e se reproduzem em fezes e podem ser ingeridos pelas pessoas quando para casa por animais ou nos calçados. As evidências demonstram que crianças pequenas que moram em casas com chãos de terra têm maior tendência de ser infectadas por parasitas intestinais, que podem causar diarreia e desnutrição, geralmente levando a problemas no desenvolvimento cognitivo ou até mesmo à morte. Chãos de cimento interrompem a transmissão de infestações parasitárias. Eles também permitem um melhor controle de temperatura e são esteticamente mais agradáveis.

Os resultados esperados informaram as perguntas da pesquisa abordadas na avaliação por Cattaneo e seus colegas. Eles esperavam concluir que substituir o chão de terra por chão de cimento reduziria a incidência de diarreia, desnutrição e deficiência de micronutrientes. Isto, por sua vez, resultaria em um melhor desenvolvimento cognitivo das crianças. Os pesquisadores também anteciparam e fizeram testes sobre melhorias no bem-estar dos adultos, conforme medido pela maior satisfação das pessoas com a situação de seu domicílio e baixas taxas de depressão e estresse.

*Fonte: Cattaneo et al. 2009.*

informações valiosas sobre como os parasitas são transmitidos e como a infestação de parasitas leva à diarreia infantil.

## A Cadeia de Resultados

Uma teoria de mudança pode ser modelada de várias formas - por exemplo, usando modelos teóricos e lógicos, estruturas lógicas, modelos de resultados e cadeias de resultados<sup>1</sup>. Tudo isso inclui os elementos básicos de uma teoria de mudança, isto é, uma cadeia causal, condições externas e influências e pressupostos-chave. Neste livro, usaremos o modelo da cadeia de resultados, pois achamos que é o modelo mais simples e mais claro para descrever a teoria de mudança no contexto operacional dos programas de desenvolvimento.

### **Conceito-chave:**

Uma cadeia de resultados estabelece a sequência de insumos, atividades e produtos que são esperados para melhorar os resultados e o resultado final.

---

Uma cadeia de resultados estabelece uma linha plausível e lógica de como uma sequência de insumos, atividades e produtos pelos quais um projeto é diretamente responsável interage com o comportamento para estabelecer trajetos através dos quais os impactos são atingidos (Figura 2.1). Ela estabelece a lógica causal desde o início do projeto, começando com os recursos disponíveis, até o final, analisando os objetivos de longo prazo. Uma cadeia básica de resultados mapeará os seguintes elementos:

*Insumos:* recursos à disposição do projeto, incluindo equipe e orçamento.

*Atividades:* medidas tomadas ou trabalho executado para converter os insumos em produtos

*Produtos:* bens e serviços tangíveis que as atividades do projeto produzem (eles ficam diretamente sob o controle do órgão responsável pela implementação.)

*Resultados:* os resultados prováveis de serem atingidos, uma vez que a população beneficiada use os produtos do projeto (geralmente, são atingidos no curto-médio prazo.)

*Resultados finais:* os objetivos finais do projeto (podem ser influenciados por vários fatores e são, tipicamente, atingidos em um período de tempo mais longo.)

A cadeia de resultados possui três partes principais:

*Implementação:* o trabalho que, planeja-se, será fornecido pelo projeto/ programa, incluindo insumos, atividades e produtos. Estas são as áreas que o órgão responsável pela implementação pode monitorar diretamente para medir o desempenho do projeto.

**Figura 2.1 O que é uma Cadeia de Resultados?**



Fonte: autores, com base em várias fontes.

*Resultados:* os resultados esperados consistem nos resultados imediatos e resultados finais, que não estão sob o controle direto do projeto e são contingentes às mudanças comportamentais dos beneficiados do programa. Em outras palavras, eles dependem das interações entre o lado da oferta (implementação) e o lado da demanda (beneficiários). Estas são as áreas sujeitas à avaliação de impacto para medir a efetividade.

*Pressupostos e riscos:* estes não estão representados na Figura 2.1. Incluem quaisquer evidências da literatura sobre a lógica causal proposta e os pressupostos dos quais ela depende, referências ao desempenho de programas semelhantes e uma menção sobre os riscos que podem afetar a obtenção dos resultados pretendidos, bem como qualquer estratégia de mitigação proposta para gerir tais riscos.

Por exemplo, imagine que o ministro da educação do país A esteja pensando em introduzir uma nova abordagem no ensino de matemática no ensino médio. Conforme apresentado na Figura 2.2, os insumos do programa

**Figura 2.2 Cadeia de Resultados de um Programa de Matemática no Ensino Médio**



Fonte: autores através de várias fontes.

incluiriam a equipe do ministério, os professores do ensino médio, um orçamento para o novo programa de matemática e as instalações municipais onde os professores de matemática serão treinados. As atividades do programa consistem na elaboração do novo currículo de matemática, além do desenvolvimento de um programa de treinamento para os professores, o próprio treinamento para os professores, o preparo, impressão e distribuição de novos livros didáticos. Os produtos são o número de professores treinados, a quantidade de livros entregues em salas de aula e a adaptação de testes padronizados ao novo currículo. Os resultados de curto prazo consistem no uso dos novos métodos e livros nas salas de aula e a aplicação dos novos testes padronizados. Os resultados de médio prazo são melhorias no desempenho dos alunos nos testes padronizados de matemática. Os resultados finais são maiores taxas de conclusão do ensino médio e taxas mais altas de empregabilidade e maiores salários para os egressos.

As cadeias de resultados são úteis para todos os projetos, independentemente de se incluirão ou não uma avaliação de impacto, pois permitem que os legisladores e gestores de programas tornem explícitos os objetivos do programa, ajudando-os, portanto, a entender a lógica causal e a sequência de eventos por trás de um programa. As cadeias de resultados também facilitam as discussões sobre o monitoramento e a avaliação, ao evidenciar quais informações precisam ser monitoradas e quais mudanças nos resultados precisam ser incluídas quando o projeto for avaliado.



Para comparar abordagens alternativas do programa, as cadeias de resultados podem ser agregadas em árvores de resultados representando todas as opções viáveis e consideradas durante o desenho do programa ou sua reestruturação. Essas árvores de resultados representam alternativas operacionais e de política para atingir objetivos específicos; elas podem ser usadas ao considerar quais opções de programas poderiam ser testadas e avaliadas. Por exemplo, se o objetivo é melhorar a educação financeira, podem ser investigadas alternativas como uma campanha publicitária versus a instrução para adultos em sala de aula.

## Hipóteses para a Avaliação

Uma vez que você tenha descrito a cadeia de resultados, as hipóteses a serem testadas usando a avaliação de impacto podem ser formuladas. No exemplo do ensino de matemática para alunos do ensino médio, as hipóteses a ser testadas poderiam ser:

- O novo currículo é superior ao antigo, no que tange a comunicação de conhecimentos de matemática.
- Professores treinados usam o novo currículo de uma forma mais efetiva que os outros professores.
- Se treinarmos os professores e distribuirmos os livros, então os professores usarão os novos livros e o currículo em sala de aula e os alunos acompanharão o currículo.
- Se treinarmos os professores e distribuirmos os livros, então os resultados dos testes de matemática aumentarão, em média, cinco pontos.
- O desempenho em matemática no ensino médio influencia as taxas de conclusão e o desempenho no mercado de trabalho.

## Selecionando Indicadores de Desempenho

Uma cadeia de resultados claramente articulada fornece um mapa útil para selecionar os indicadores que serão medidos ao longo da cadeia. Eles incluirão indicadores usados tanto para monitorar a implementação do programa quanto para avaliar os resultados. Mais uma vez, é útil envolver as partes interessadas no programa na seleção destes indicadores, para garantir que os indicadores selecionados sejam boas medidas do desempenho

### ***Conceito-chave:***

Bons indicadores são específicos, mensuráveis, atribuíveis, realistas e direcionados (tem foco).

do programa. O acrônimo *SMART* é amplamente usado como regra útil e geral para garantir que o indicador usado seja:

- *Específico*: para medir a informação que, de fato, se quer medir.
- *Mensurável*: para garantir que as informações sejam, de fato, obtidas.
- *Atribuível*: para garantir que cada medida seja relacionada aos esforços do projeto.
- *Realista*: para garantir que os dados possam ser obtidos em tempo adequado, com frequência razoável e a um custo também razoável.
- *Tem foco (direcionado)*: a população-alvo.

Ao escolher indicadores, lembre-se de que é importante identificar indicadores em toda a cadeia de resultados e não somente nos resultados finais, de modo que você possa rastrear a lógica causal de qualquer resultado observado no programa. Mesmo quando você realiza uma avaliação de impacto, é importante rastrear indicadores de implementação, para que você possa determinar se as intervenções estão sendo executadas conforme planejadas, se atingiram os beneficiados pretendidos e se chegaram a tempo (vide Kusek e Rist 2004 ou Imas e Rist 2009 para discussões de como selecionar indicadores de desempenho). Sem estes indicadores ao longo da cadeia de resultados, a avaliação de impacto produzirá somente uma “caixa preta” que identifica se os resultados previstos se materializaram ou não, mas não será capaz de explicar a razão do sucesso ou do fracasso.

Além de selecionar os indicadores, é útil também considerar os arranjos para a produção de dados. A Tabela 2.1 lista os elementos básicos de um plano de monitoramento e avaliação (M&A), cobrindo os arranjos necessários para produzir cada um dos indicadores de modo confiável e tempestivo.

## **Roteiro para as Partes 2 e 3**

Nesta primeira parte do livro, discutimos por que uma avaliação de impacto deveria ser realizada e quando vale a pena realizá-la. Revisamos os vários objetivos que uma avaliação de impacto pode atingir e destacamos as questões de políticas fundamentais que uma avaliação pode abordar. Insistimos na necessidade de se traçar cuidadosamente a teoria de mudança que explica os canais por meio dos quais um programa pode influenciar os resultados finais. Essencialmente, as avaliações de impacto testam se, na prática, a teoria de mudança funciona ou não.

**Tabela 2.1 Elementos de um Plano de Monitoramento e Avaliação**

Elemento	Descrição
Resultados esperados (resultados imediatos e produtos)	Obtidos dos documentos de elaboração do programa e da cadeia de resultados.
Indicadores (com linhas de base e metas indicativas)	Derivados da cadeia de resultados; os indicadores devem ser <i>SMART</i> .
Dados das fontes	Fonte e localização de onde os dados serão obtidos - por exemplo, uma pesquisa, uma revisão, uma reunião de atores relevantes.
Frequência de dados	Frequência de disponibilidade de dados.
Responsabilidades	Quem é responsável pela organização da coleta de dados e a verificação da qualidade e da fonte dos dados?
Análise e relatório	Frequência de análise, método de análise e responsabilidade pelo relatório.
Recursos	Estimativa de recursos requeridos e comprometidos com a execução das atividades M&A planejadas de M&A.
Uso final	Quem receberá e revisará as informações? A que propósito elas servem?
Riscos	Quais são os riscos e pressupostos na execução das atividades M&A planejadas de M&A? Como eles poderiam afetar os eventos de M&A e a qualidade dos dados?

Fonte: Adaptado do UNDP 2009.

Na Parte 2, consideramos *como avaliar*, revisando várias metodologias alternativas que produzem grupos de comparação válidos e permitem a estimativa de impactos válidos do programa. Começamos a introduzir o *cenário contrafactual* como o cerne de qualquer avaliação de impacto, detalhando as propriedades que a estimativa do cenário contrafactual deve ter e fornecendo exemplos de estimativas inválidas ou falsas do mesmo. Apresentamos, então, opções de avaliação de impacto que podem produzir estimativas válidas do contrafactual. Em particular, discutimos a intuição básica por trás de quatro categorias de metodologias: métodos de *seleção aleatória*, *desenho de regressão descontínua*, *diferenças-em-diferenças* e *método de pareamento*. Discutimos por que e como cada método pode produzir uma estimativa válida do contrafactual, em que contexto de política cada um pode ser usado e as principais limitações de cada método. Ao longo desta parte do livro, um estudo de caso - o Programa de Subsídio ao Seguro Saúde<sup>2</sup> - é usado para

ilustrar como os métodos podem ser aplicados. Além disso, apresentamos exemplos específicos de avaliações de impacto que fizeram uso dos diferentes métodos.

A Parte 3 delinea os passos para implementar, administrar ou contratar uma avaliação de impacto. Presumimos que, a essa altura, os objetivos da avaliação já tenham sido definidos, a teoria de mudança tenha sido formulada e as perguntas da avaliação tenham sido especificadas. Revisamos as questões-chave que precisam ser respondidas ao formular-se um plano de avaliação de impacto. Começamos fornecendo regras claras para decidir de onde virão os grupos de comparação. É estabelecida uma estrutura simples para determinar quais das metodologias de avaliação de impacto apresentadas na Parte 2 são as mais adequadas para um determinado programa, dependendo de suas regras operacionais. Revisamos, então, os passos nas quatro fases de realização de uma avaliação: operacionalizar a avaliação planejada, escolher uma amostra, coletar dados e produzir e disseminar os resultados.

## Notas

1. A University of Wisconsin-Extension (2010) contém uma discussão detalhada sobre como construir uma cadeia de resultados, assim como uma lista abrangente de referências. Imas e Rist (2009) fornecem uma boa revisão de teorias de mudança.
2. NOTA DO TRADUTOR: o termo “seguro saúde” não é comum no Brasil devido à existência do Sistema Único de Saúde (SUS), que oferece serviços de saúde gratuitos à população carente.

## Referências

- Cattaneo, M., Galiani, Gertler, Martinez e Titiunik. 2009. Housing, Health and Happiness. *American Economic Journal: Economic Policy* 1 (1): 75–105.
- Imas, L. & Rist, R. (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: Banco Mundial.
- Kusek, Z. & Rist. (2004). *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington, DC: Banco Mundial.
- PNUD, Programa das Nações Unidas para o Desenvolvimento. (2009). *Handbook on Planning, Monitoring and Evaluating for Development Results*. Nova York: PNUD.
- Universidade de Wisconsin-Extensão. (2010). Enhancing Program Performance with Logic Models. Curso on-line. Disponível em: <http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html>.