SÉRIE Avaliação na Prática

Lycia Lima | André Portela

GUIA CLEAR

MONITORAMENTO E AVALIAÇÃO DE POLÍTICAS PÚBLICAS

Do diagnóstico à decisão

clear avaliação áfrica lusófona e brasil

FGV EESP ESCOLA DE ECONOMIA DE SÃO PAULO

global evaluation initiative

© FGV CLEAR 2025

Autores:
Lycia Lima, professora da FGV EESP e pesquisadora do FGV CLEAR
André Portela Souza, professor da FGV EESP e diretor do FGV CLEAR

Colaboradores:
Alei Fernandes Santos, Amanda Cappellazzo Arabage, Amanda Schutze, Ana Carolina Marinato de Resende, Beatriz Burattini, Bernardo Ostrovski, Bruno Guimarães de Melo, Caio de Souza Castro, Camila Mata Machado Soares, Camila Magalhaes, Cecilia Florentino, Fabrícia Volotão Peixoto, Fernando Gonçalves Marques, Gabriel Weber, Gabriela Lacerda, Gabriela Serpa, Giovana Pereira, Giovanna Ribeiro, Gustavo Nebó Garcia, Lara Mesquita Ramos, Leandro Lima, Lorena Mello e Figueiredo, Lucas Finamor, Lucas Gerez, Luan Paciencia, Marilia Firmiano, Marina Lafer, Marina Ribeiro, Michel Szklo, Nicole Mourad Pereira, Patrícia Franco Ravaioli, Pedro Molina Ogeda, Priscilla Bacalhau, Rocio Lorenzo, Ronan Cunha, Victor Simões Dornelas e Vitor Menezes.

Citação recomendada:
Lima, L., Souza, A. P. (2025) Guia Clear monitoramento e avaliação de políticas públicas: do diagnóstico à decisão. São Paulo, FGV CLEAR.

Apresentação

O Centro CLEAR

Este guia foi desenvolvido pelo **Centro FGV CLEAR da Fundação Getulio Vargas, sediado na Escola de Economia de São Paulo (FGV EESP)** desde 2015. O FGV CLEAR é o centro regional da Iniciativa CLEAR (Centers for Learning on Evaluation and Results) para o Brasil e os países lusófonos, atuando para fortalecer a cultura de avaliação e a gestão informada por evidências.

Desde 2020, a Iniciativa CLEAR integra a **Global Evaluation Initiative (GEI)**, uma coalizão liderada pelo Banco Mundial e pelo Programa das Nações Unidas para o Desenvolvimento (PNUD), que busca promover o uso sistemático de avaliação e fortalecer os ecossistemas nacionais de Monitoramento e Avaliação (M&A) em países em desenvolvimento.

O FGV CLEAR atua com governos, organizações multilaterais, sociedade civil e academia, oferecendo capacitação, assistência técnica, geração e disseminação de conhecimento científico na área.

O conteúdo desta publicação reflete a experiência acumulada ao longo de mais de uma década de atuação do centro, com projetos desenvolvidos em diversos contextos no Brasil e na África Lusófona, e visa contribuir para o uso qualificado de evidências na formulação, implementação e aprimoramento de políticas públicas.

Apresentação dos autores

Os autores da presente publicação são os **professores Lycia Lima e André Portela**, docentes de políticas públicas da Escola de Economia de São Paulo da Fundação Getulio Vargas (FGV EESP) e **cofundadores do Centro FGV CLEAR para o Brasil e a África Lusófona**. Este guia foi concebido a partir da experiência acumulada pelo centro ao longo da última década na avaliação de políticas públicas, sob a liderança de ambos os professores.

A equipe técnica do FGV CLEAR teve participação fundamental na consolidação dessa abordagem ao longo da última década. Os nomes dos colaboradores estão listados na ficha técnica.

Lycia Lima é doutora em Administração Pública e Governo pela FGV, mestre em Economia do Desenvolvimento pela Universidade de Londres (SOAS) e graduada em Economia pela Universidade Federal de Minas Gerais. Com atuação consolidada no desenho e avaliação de políticas públicas, Lycia desenvolveu sua carreira em organizações internacionais, governos e academia. É autora de diversos artigos acadêmicos, capítulos de livros e publicações técnicas na área de avaliação de políticas públicas, com circulação nacional e internacional. Na FGV EESP, leciona

regularmente disciplinas de graduação e pós-graduação, como desenho e avaliação de políticas públicas, microeconomia, econometria e tópicos em políticas públicas. Esteve à frente do FGV CLEAR, ao lado do professor André Portela, durante os dez primeiros anos do centro e, atualmente, segue como pesquisadora do CLEAR e pesquisadora associada dos centros FGV Cidades e FGV Clima.

André Portela Souza é doutor em Economia pela Universidade Cornell, mestre pela Universidade de São Paulo (USP) e graduado pela Universidade Federal da Bahia. É professor titular de políticas públicas da FGV EESP, onde ministra disciplinas como Microeconomia do Desenvolvimento, Microeconomia Aplicada ao Brasil e Econometria. É diretor do FGV CLEAR e coordenador do Centro de Microeconomia Aplicada (C-Micro/FGV EESP). Foi professor da Universidade de São Paulo, professor associado adjunto na Universidade Vanderbilt e professor visitante na Universidade Cornell. É autor de diversos artigos acadêmicos publicados em periódicos científicos nacionais e internacionais.

Mais do que a produção de um único grupo, este trabalho reflete o aprendizado contínuo proporcionado pelas inúmeras interações com parceiros da **Global Evaluation Initiative (GEI)**, governos, organizações da sociedade civil, instituições acadêmicas e alunos — todos fundamentais para o aprimoramento das práticas e metodologias aqui apresentadas.

Sobre o guia

Este guia se destina a todas as pessoas e instituições interessadas em aprimorar políticas públicas e programas sociais[1] por meio do uso qualificado de evidências. Seu público-alvo inclui gestores e gestoras de todas as esferas de governo, membros do terceiro setor e do setor privado, consultores, pesquisadores, estudantes de graduação e pós-graduação e cidadãos comprometidos com a busca pelos melhores resultados em suas políticas e programas.

Esta publicação apresenta a abordagem adotada pelo Centro FGV CLEAR ao longo de sua primeira década de atuação, com base em experiências concretas no Brasil e na África Lusófona. Essa abordagem já foi aplicada em contextos diversos — incluindo governos nacionais, estaduais e municipais — e busca fortalecer práticas de gestão baseada em evidências por meio da institucionalização da prática avaliativa.

Embora o foco esteja em políticas públicas, os conceitos, ferramentas e etapas aqui apresentados também podem ser adaptados por organizações da sociedade civil, especialmente aquelas envolvidas no desenvolvimento e gestão de programas sociais. Trata-se, portanto, de um conteúdo aplicável a diferentes setores comprometidos com o uso qualificado de evidências na tomada de decisão.

Acreditamos que a avaliação é uma ferramenta importante para a gestão, capaz de

[1] Embora políticas públicas e programas sociais tenham distinções conceituais, as ferramentas de avaliação apresentadas neste guia são aplicáveis a ambos. Por esse motivo, usamos os dois termos ao longo do texto, respeitando suas especificidades, mas com foco na aplicação prática do conteúdo.

apoiar o trabalho cotidiano do gestor público ao fornecer informações relevantes e robustas para a tomada de decisão sobre seus programas.

Adotamos o conceito de ciclo da política pública como ponto de partida, entendido como uma sequência de etapas interligadas: identificação do problema, formulação, implementação, monitoramento, avaliação e tomada de decisão. Embora na prática essas etapas nem sempre ocorram de forma linear, elas fornecem uma estrutura útil para organizar o raciocínio avaliativo e orientar a gestão orientada para resultados.

Ainda que o ciclo em sua versão original preveja uma fase dedicada à avaliação, a prática avaliativa deve estar presente em todas as etapas — inclusive antes da criação de uma política ou programa. Incorporar esse olhar desde o início permite qualificar o diagnóstico do problema público, identificar suas causas e efeitos, compreender os grupos atingidos e os contextos envolvidos. A partir de um diagnóstico preciso, é possível planejar uma política com base na análise de diferentes alternativas, desenhando, assim, um programa com maior potencial de alcançar os resultados pretendidos ao menor custo possível. Essa é a essência da análise ex ante, que orienta o desenho de políticas com maior coerência interna e potencial de impacto.

Durante a implementação, o monitoramento contínuo e a avaliação são cruciais para verificar se a política está sendo executada conforme o planejado e corrigir desvios de rota quando necessário. Já a análise ex post é realizada após pelo menos um ciclo de implementação, com o objetivo de verificar se os resultados e impactos esperados foram alcançados — e a que custo. Essas avaliações subsidiam decisões sobre o futuro da política: manter, expandir, reformular ou interromper.

Ao longo do guia, apresentamos ferramentas e métodos que apoiam o uso de evidências desde o diagnóstico até a decisão final sobre a política. Nosso ponto de partida é o entendimento de que a avaliação não é uma etapa isolada ao final de um programa, mas uma forma de pensar e agir que deve permear toda a trajetória da política pública.

O campo da avaliação é amplo, interdisciplinar e diverso. Ele reúne contribuições da ciência política, economia, administração, sociologia, estatística, antropologia, entre outras áreas. Diante da diversidade de abordagens teóricas e metodológicas existentes, este guia não busca esgotar o tema. Seu objetivo é oferecer um framework prático e acessível, que apoie equipes e gestores públicos no desenvolvimento de competências avaliativas e na tomada de decisões mais qualificadas. Ao longo do texto, indicamos referências e fontes para quem desejar se aprofundar em conceitos, métodos ou abordagens específicas não abordadas de forma exaustiva aqui.

O guia está dividido em quatro partes que se complementam e mostram como os diferentes tipos de avaliação devem estar presentes **ao longo de todo o ciclo da política pública**. A primeira parte apresenta o papel da avaliação como elemento transversal à formulação, implementação e revisão das políticas. As partes centrais aprofundam os dois principais momentos de aplicação da avaliação: antes da implementação (análise ex ante) e após a execução (análise ex post). Por fim, a

última parte discute como transformar evidências em decisões, destacando estratégias para promover o uso efetivo da avaliação na gestão pública.

Parte I – Do diagnóstico à decisão: a avaliação no ciclo da política pública

Apresenta o papel da avaliação ao longo de todas as fases do ciclo de políticas públicas, com base na lógica da geração e do uso de evidências para qualificar as decisões. Estabelece os fundamentos conceituais que orientam o guia e mostra como integrar avaliação de forma transversal ao ciclo de uma política pública.

Parte II – Análise ex ante: planejamento informado por evidências

Concentra-se nas etapas anteriores à implementação da política, como o diagnóstico do problema, a análise de causas e efeitos, a definição de grupos-alvo, o levantamento de evidências, a avaliação de alternativas e a formulação da política com base na Teoria da Mudança (TdM). Também introduz os principais conceitos para a elaboração de um plano de monitoramento e avaliação. Esta seção orienta o desenho de políticas públicas mais coerentes, eficazes e com maior potencial de impacto.

Parte III – análise ex post: monitorar, aprender e aprimorar

Reúne métodos e abordagens aplicados após a implementação da política ou programa. Engloba o monitoramento e a avaliação da implementação, que analisam se a política foi executada conforme o planejado, com cobertura e qualidade adequadas. Inclui também a apresentação da avaliação de resultados, de impacto e de custo-benefício, que buscam mensurar os efeitos gerados e seu valor para a sociedade. Apresenta, ainda, a avaliação executiva, uma modalidade de análise rápida que considera múltiplas dimensões da política para apoiar decisões em contextos com prazos curtos. A seção discute como essas abordagens podem apoiar o aperfeiçoamento de políticas em curso e embasar decisões sobre sua continuidade, expansão ou reformulação.

Parte IV – Da evidência à decisão: reflexões para a tomada de decisão

Esta parte oferece orientações práticas e reflexões estratégicas sobre como transformar evidências em decisões qualificadas sobre políticas públicas. Com base nas etapas anteriores do guia, discute como as informações geradas ao longo do ciclo — desde o diagnóstico até as avaliações ex post — podem embasar decisões sobre continuidade, reformulação, expansão ou encerramento de políticas. Aborda também os fatores que influenciam o uso efetivo das evidências, como a viabilidade política, a mobilização de atores, os valores sociais e as janelas de oportunidade.

Nosso convite ao longo deste guia é para que a avaliação seja compreendida não como um fim em si mesma, mas como uma forma de pensar — uma ferramenta poderosa de gestão, aprendizado institucional e aprimoramento contínuo das políticas públicas. Acreditamos que o uso consciente de evidências contribui para decisões mais acertadas e para o fortalecimento da democracia, por meio da transparência, da responsabilização e do diálogo com a sociedade.

Este é um guia pensado para apoiar o aprendizado de gestores públicos, pesquisadores, avaliadores e estudantes de graduação e pós-graduação interessados em

desenho e avaliação de políticas públicas. Independentemente da formação ou área de atuação, ele oferece um ponto de partida para pensar a avaliação de forma prática, crítica e conectada ao cotidiano de quem atua com políticas públicas e programas sociais.

Esta publicação marca o início da série "Avaliação na Prática", lançada pelo FGV CLEAR com o objetivo de promover o uso de evidências na gestão pública[2].

Este primeiro volume apresenta uma visão abrangente da avaliação de políticas públicas, com foco especial em sua integração ao ciclo da política. Mais do que uma introdução à área, o guia sistematiza conceitos, etapas e práticas essenciais para a implementação de uma gestão informada por evidências. Ao longo da série Avaliação na Prática, novos títulos trarão aprofundamentos em métodos e abordagens específicas, complementando e expandindo os temas aqui tratados.

Convidamos você a seguir conosco nessa jornada, explorando as publicações disponíveis no site do FGV CLEAR e acessando nossos cursos online gratuitos, oferecidos pela plataforma de educação executiva da FGV. Saiba mais em www.fgvclear.org.

[2]Esta publicação integra a série Avaliação na Prática. A série reúne guias técnicos voltados a gestores públicos, avaliadores e formuladores de políticas e programas, com foco na aplicação prática de métodos e ferramentas de avaliação.

Conheça os demais guias da série, que exploram distintas abordagens e temas:

- Avaliação de Impacto
- Avaliação de Impacto: método de Diferença em Diferenças (DiD)
- Cálculo de Poder Estatístico
- Análise de Custo-Benefício
- Avaliação ex-ante de Impacto: Modelagem de Microssimulação

Novos volumes serão publicados periodicamente e estarão disponíveis no site do FGV CLEAR: www.fgvclear.org.

Sumário

Apresentação 3

Parte I – Do diagnóstico à decisão: a avaliação no ciclo da política pública 16

1 Avaliação como elemento central da gestão pública informada por evidências 17

2 O Ciclo da Política Pública 19

3 Avaliação ao longo do Ciclo da Política Pública 21

3.1 Análise ex ante: antes da implementação da política 22

3.2 Análise ex post: compreendendo os efeitos da política após sua implementação 23

3.3 Análise ex ante ou ex post? Diferentes, complementares, essenciais 23

3.4 Avaliação executiva: uma ferramenta transversal e ágil 25

4 Etapas do ciclo e o uso de avaliações em cada fase 26

4.1 Identificação do problema 26

4.2 Formulação da política 26

4.3 Implementação 27

4.4 Avaliação de resultados, impactos e custos 27

4.5 Tomada de decisão 28

5 Redesenhando o ciclo: avaliação como eixo transversal 29

6 A abordagem do ciclo e a importância da institucionalização da avaliação nos governos de decisão 31

7 Conclusão 33

Parte II – Análise ex ante: planejamento informado por evi-

dências **35**

1 Identificação e diagnóstico do problema público **36**

1.1 Por que é importante detalhar bem o problema? 36

1.2 Etapas para a identificação e diagnóstico de um problema público . 37

1.2.1 Delimitação do problema 37

1.2.2 Levantamento de evidências 40

1.2.3 Análise da distribuição do problema 43

1.2.4 Avaliação da relevância do problema e priorização 45

1.2.5 Mapeamento de causas e consequências: a árvore do problema como ferramenta de diagnóstico 48

2 Formulação da Política: escolhas estratégicas para enfrentar o problema **55**

2.1 Buscando alternativas de políticas com base em evidências 55

2.1.1 Ponto de Partida: Delimitando o público da política: população potencial, objetivo e beneficiária 56

2.1.2 Onde encontrar boas evidências? 57

2.1.3 Atenção à qualidade da evidência 59

2.1.4 Aprender também com o que não funcionou 61

2.1.5 E se a política for nova? 62

2.1.6 Quando há várias alternativas possíveis: como escolher? . 62

3 Estruturando a lógica da política: a Teoria do Programa e a Teoria da Mudança **67**

3.1 O que é a Teoria da Mudança? 68

3.2 Os componentes da Teoria da Mudança 68

3.2.1 Insumos: O que é necessário mobilizar para que a política exista? 69

3.2.2 Atividades: O que será feito com os insumos mobilizados? . 69

3.2.3 Produtos: O que será entregue diretamente ao público-alvo? 69

3.2.4 Resultados: O que deve mudar no público-alvo em decorrência do uso dos produtos? 70

3.2.5 Impactos: Quais são as transformações estruturais esperadas no longo prazo? 70

3.2.6 Hipóteses e riscos: os elos invisíveis da cadeia 71

3.3 Como construir a Teoria da Mudança na prática? 73
3.4 Representação visual: o diagrama da Teoria da Mudança 76
3.5 Identificando falhas: teoria ou implementação? 78
3.5.1 Falha de implementação 79
3.5.2 Falha na teoria (ou falha no desenho) 79

4 Estratégias de Mitigação de Riscos: antecipar problemas para proteger os resultados 81
4.1 Como os riscos se conectam à avaliação? 81
4.2 Etapas para elaborar estratégias de mitigação 82

5 Avaliação de Desenho: Análise Crítica da Teoria da Mudança 84
5.1 O que é a avaliação de desenho? 84
5.2 Quando aplicar a avaliação de desenho? 84
5.3 Dimensões centrais da avaliação de desenho 85

6 Plano de monitoramento e avaliação: usar a Teoria da Mudança para transformar estratégia em evidência 88
6.1 Plano de Monitoramento: acompanhar para aprender, ajustar e entregar resultados 89
6.1.1 Como a Teoria da Mudança orienta o monitoramento? . . . 89
6.1.2 Componentes de um bom plano de monitoramento 90
6.2 Plano de Avaliação: testar hipóteses, verificar riscos e aprender com os resultados 92
6.2.1 A Teoria da Mudança como geradora de perguntas avaliativas 93
6.2.2 Tipologia de perguntas avaliativas segundo a Teoria da Mudança 93
6.2.3 Critérios analíticos da OCDE como lentes complementares 95
6.2.4 Como integrar critérios da OCDE nos diferentes tipos de avaliação? 97
6.2.5 Outras abordagens para enriquecer o plano de avaliação: Avaliação com foco em equidade e gênero 99
6.2.6 Outras abordagens para enriquecer o plano de avaliação: Avaliação participativa 100

6.2.7 Outras abordagens para enriquecer o plano de avaliação: *Footprint Evaluation* e justiça climática como lentes transversais 100
6.2.8 Outras abordagens para enriquecer o plano de avaliação: Avaliação de mudanças transformadoras 102
6.2.9 Estruturando o plano de avaliação 103

7 Conclusão **105**

Parte III – Análise ex post: analisar, aprender e aprimorar 107

1 Monitoramento: colocando o plano em prática **110**
1.1 O que o monitoramento permite observar? 111
1.2 Monitoramento orientado ao uso da informação 114
1.3 Monitoramento e transparência 115

2 Avaliação de Implementação e Processos de Políticas Públicas **117**
2.1 Principais Perguntas Avaliativas 119
2.2 Métodos e Abordagens: A Importância dos Métodos Mistos 120
2.2.1 Fontes de Dados Quantitativos 120
2.2.2 Fontes de Dados Qualitativos 121
2.2.3 Triangulação e Abordagem Integrada 122
2.3 Etapas para a avaliação de processos 123
2.4 Armadilhas a evitar 125
2.4.1 Avaliação de processos ou avaliação de resultados? 125
2.4.2 Fazer julgamentos precipitados sem considerar o contexto 125
2.4.3 Sobrecarregar a coleta de dados 126
2.4.4 Ignorar vieses de avaliadores e informantes 126
2.4.5 Não planejar o uso dos resultados 126

3 Introdução à Avaliação de Resultados **129**
3.1 Finalidades para a Avaliação de Resultados 130
3.2 Avaliação de resultados versus outros tipos de avaliação 131
3.2.1 Avaliação de Resultados e Avaliação de Impacto 131
3.2.2 Avaliação de Resultados vs. Avaliação de Processos 132

3.2.3 Avaliação de Resultados vs. Monitoramento 133
3.3 Principais Perguntas Avaliativas 134
3.4 Métodos e Abordagens para Avaliação de Resultados 135
3.4.1 Fontes e tipos de dados 135
3.4.2 Métodos mais utilizados na avaliação de resultados 137
3.4.3 Outras abordagens qualitativas úteis para contextos complexos 138
3.4.4 Combinação de métodos: triangulação e coerência com a Teoria da Mudança 138

4 Avaliação de Impacto de Políticas Públicas 142
4.1 Entendendo a lógica da causalidade: o desafio do contrafactual . . 143
4.1.1 Que efeito estamos medindo? Efeito Médio do Tratamento, Efeito Médio sobre os Tratados, Efeito Médio Local do Tratamento e outras estimativas possíveis 145
4.2 Métodos Quantitativos de Avaliação de Impacto 146
4.2.1 Experimentos aleatórios (Randomized Controlled Trials – RCTs) 146
4.2.2 Regressão Descontínua (Regression Discontinuity Design – RDD) 147
4.2.3 Variáveis Instrumentais (Instrumental Variables – IV) . . . 148
4.2.4 Diferenças em Diferenças (Difference-in-Differences – DiD) 150
4.2.5 Controle Sintético 151
4.2.6 Pareamento 152
4.3 Como interpretar os resultados? 154
4.3.1 Entendendo o valor estimado 154
4.3.2 Significância estatística e intervalos de confiança 155
4.3.3 Direção, magnitude e relevância prática 155
4.3.4 Causalidade e linguagem apropriada 156
4.3.5 Efeitos heterogêneos 156
4.4 Como escolher o método mais adequado 157
4.4.1 Que tipo de pergunta você quer responder? 157
4.4.2 O quanto os grupos são diferentes entre si? (e o que dá para observar?) 158
4.4.3 Que tipo de dados você tem (ou pode ter)? 158

4.4.4 Como escolher o método mais adequado na prática? 159
4.5 Dados qualitativos e métodos mistos: por que são importantes na avaliação de impacto? 160
4.6 Desenho e implementação da avaliação de impacto 161

5 Avaliação de Custo-Benefício e Custo-Efetividade 170
5.1 Por que avaliar economicamente uma política? 170
5.2 ACE e ACB: qual a diferença? 171
5.2.1 O que a ACE responde? 171
5.2.2 O que a ACB responde? 171
5.3 Quando usar cada abordagem? 172
5.4 Métricas principais da avaliação econômica 173
5.4.1 Custo por unidade de resultado (RCE) 174
5.4.2 Razão Benefício-Custo (RBC) 175
5.4.3 Valor Presente Líquido (VPL) 175
5.5 Como interpretar essas métricas? 176
5.6 Etapas de uma avaliação econômica 176
5.6.1 Definindo a política e a população avaliada 176
5.6.2 Mapeando os custos da política 177
5.6.3 Estimando os efeitos da política — com base em avaliação de impacto 177
5.6.4 Monetizando os efeitos 178
5.6.5 Cálculo da Métrica da Política 178
5.6.6 Análise de sensibilidade 179
5.6.7 Interprete os achados com foco na decisão 179
5.7 Limitações e desafios práticos 180

6 Avaliação Executiva 184
6.1 Situações em que a Avaliação Executiva é Recomendada 185
6.2 Limites de Aplicação da Avaliação Executiva 186
6.3 Etapas de uma Avaliação Executiva 189
6.3.1 Definição do escopo e das perguntas avaliativas 190
6.3.2 Coleta e análise rápida de evidências 190
6.3.3 Síntese dos resultados e comunicação para a utilidade . . . 192

6.4 Estrutura recomendada para o relatório de avaliação executiva . . 193
6.4.1 Caracterização da política avaliada 193
6.4.2 Perguntas avaliativas e critérios de análise 194
6.4.3 Métodos e fontes de dados 194
6.4.4 Achados e análise por dimensão 194
6.4.5 Conclusões e recomendações 195
6.4.6 Anexos 196

7 Conclusão **202**

Parte IV – Tomada de decisão 204

1 A Importância da Geração e Uso de Evidências **206**

2 A Qualidade das Evidências **207**

3 Valores, raciocínio e evidências nas tomadas de decisão **209**
3.1 A Importância do modelo teórico (mental) adequado 209
3.2 Evidências, Incertezas e Critérios de Decisão 210
3.2.1 Decidir na prática: evidências como guia, não como sentença 213
3.2.2 Critérios práticos que influenciam a decisão 213
3.2.3 Caminhos possíveis após a avaliação: manter, expandir, ajustar ou encerrar 214

4 O Contexto Político **216**

5 Conclusão **219**

Referências **220**

Apêndice I - Ferramentas baseadas em IA para apoiar a busca e síntese de evidências **231**

PARTE I

Do diagnóstico à decisão: a avaliação no ciclo da política pública

Parte I – Do diagnóstico à decisão: a avaliação no ciclo da política pública[3]

[3] A Parte I deste guia se baseia, de forma geral, na literatura técnica produzida por OCDE (2020), Secchi (2014), Kraft e Furlong (2019), FGV CLEAR (2025), Kusek e Rist (2004), e Parkhurst (2017).

1 Avaliação como elemento central da gestão pública informada por evidências

A gestão pública informada por evidências configura um princípio orientador cada vez mais presente nos governos: o de que decisões sobre políticas públicas devem se apoiar, sempre que possível, em informações qualificadas sobre o problema, os contextos, as alternativas disponíveis e seus possíveis efeitos. Mais do que uma agenda técnica, trata-se de uma diretriz de boa governança.

Essa perspectiva reconhece o potencial das evidências para aprimorar políticas públicas em todas as etapas do seu ciclo de vida — da identificação do problema até a decisão sobre sua continuidade ou reformulação. Também promove uma cultura organizacional que valoriza o aprendizado, a transparência e o uso estratégico de informações para gerar resultados concretos.

Nesse contexto, **o monitoramento e a avaliação de políticas públicas se destacam como instrumentos fundamentais para a produção de evidências robustas e úteis à gestão**. O monitoramento consiste no acompanhamento contínuo e sistemático da implementação da política, com foco no progresso, na cobertura e na qualidade das entregas em relação ao que foi planejado. Já a avaliação busca analisar, de forma estruturada e com métodos apropriados, a lógica de intervenção, os processos, os resultados e os impactos gerados pela política. Juntos, monitoramento e avaliação permitem compreender o que está funcionando, por que está funcionando (ou não), e quais ajustes são necessários para alcançar os melhores resultados.

No entanto, evidências não possuem um valor intrínseco. Elas apenas assumem uma potência real quando utilizadas de forma sistemática e institucionalizada, como parte de uma governança estruturada. Por isso, diversos países criaram sistemas nacionais de avaliação e uso de evidências, muitas vezes respaldados por marcos legais, instâncias de coordenação, protocolos técnicos e engajamento de múltiplos atores. Casos como os do Brasil, Chile, México, Reino Unido e África do Sul ilustram arranjos institucionais diversos, mas dotados de uma mesma lógica: garantir que o conhecimento produzido seja efetivamente incorporado à gestão pública (OCDE, 2020).

No Brasil, já há legislações e iniciativas voltadas à avaliação de políticas públicas em vários estados e no Governo Federal. O desafio atual está em consolidar práticas duradouras que conectem avaliação, planejamento e tomada de decisão. Isso requer investimentos em capacidades técnicas, liderança política, valorização da aprendizagem e produção contínua de conhecimento relevante e acessível.

Ao longo deste guia, apresentamos uma abordagem integrada para fortalecer esse tipo de gestão: a avaliação como instrumento transversal ao ciclo da política pública, colocando as evidências a serviço do aprendizado, da melhoria contínua e de decisões mais bem fundamentadas. Avaliar é compreender, formular hipóteses, testar caminhos e ajustar rotas.

2 O Ciclo da Política Pública

O ciclo da política pública é uma ferramenta conceitual amplamente adotada por governos [4], organismos internacionais e centros de pesquisa para orientar o desenho, a gestão e a avaliação de políticas públicas. Ele parte da ideia de que, mesmo em contextos complexos e dinâmicos, é possível organizar o processo de formulação de políticas em uma sequência lógica de etapas, que ajudam a estruturar o raciocínio e a tomada de decisão (SECCHI, 2014; KRAFT & FURLONG, 2019).

Neste guia, usamos como ponto de partida uma versão do ciclo inspirada em autores como Secchi (2014) e Kraft e Furlong (2019), que organizam esse processo em cinco grandes etapas:

1. Identificação do problema
2. Formulação da política
3. Implementação
4. Avaliação
5. Tomada de decisão

Essa estrutura é representada na **Figura 1**:

Figura 1: O ciclo da política pública



Fonte: Elaborado com base em Secchi (2014) e Kraft e Furlong (2019).

[4] O ciclo da política pública é utilizado em guias para a realização de avaliações no Reino Unido (2022) e na Colômbia (2024), por exemplo.

Embora, na prática, essas etapas nem sempre ocorram de forma linear ou perfeitamente ordenada — muitas vezes se sobrepondo ou se retroalimentando —, o ciclo oferece uma referência útil para pensar as políticas públicas de forma estruturada, especialmente quando se busca incorporar o uso de evidências ao longo dos processos de tomada de decisão.

Um ponto essencial da abordagem apresentada neste guia é que a avaliação não deve ser tratada como uma etapa isolada do ciclo. Tradicionalmente representada como algo que acontece apenas "depois" da política — para medir se ela funcionou ou não —, a avaliação pode oferecer muito mais. Propomos aqui um olhar diferente: a prática avaliativa deve estar presente desde o início, integrada a todas as fases do ciclo da política pública — tanto antes da implementação (ex ante) quanto depois (ex post), como discutido a seguir.

Incluir a avaliação desde as etapas iniciais permite qualificar o diagnóstico do problema público, compreender melhor seus determinantes, conhecer as alternativas possíveis, construir uma Teoria da Mudança (TdM) — isto é, uma representação lógica de como e por que a política deve funcionar — e preparar sua implementação para que possa ser monitorada e, posteriormente, avaliada em relação aos seus resultados e impactos.

3 Avaliação ao longo do Ciclo da Política Pública

Ao longo dos últimos anos, consolidou-se a compreensão de que políticas públicas mais eficazes são aquelas que incorporam o uso sistemático de evidências em todas as suas etapas. Isso significa pensar na avaliação não como algo que se faz "ao final", mas como uma **lógica transversal** que acompanha a política desde antes da sua criação.

Nesse sentido, distinguimos dois grandes momentos em que as ferramentas avaliativas podem (e devem) ser mobilizadas: o momento ex ante e o momento ex post, conforme representado na **Figura 2**. Embora o modelo tradicional do ciclo da política pública apresente a avaliação como uma fase posterior à implementação, reforçamos a ideia de que ela deve estar presente em todas as etapas — como discutiremos com mais detalhes ao longo deste guia.

Figura 2: Avaliação ao longo do ciclo: ex ante e ex post



Fonte: Elaborado com base em Secchi (2014) e Kraft e Furlong (2019).

3.1 Análise ex ante: antes da implementação da política

As análises ex ante são realizadas **antes do lançamento de uma política pública**. Seu papel é contribuir para que a política seja **bem desenhada desde o início**, com base em evidências sobre o problema a ser enfrentado, as alternativas disponíveis e o contexto em que será implementada.

Durante o diagnóstico, aplicamos a**valiações diagnósticas** ou **de necessidades**, que nos ajudam a compreender profundamente o problema – suas causas, consequências, dinâmicas e quais são os grupos mais afetados. Esse passo é essencial para evitar diagnósticos superficiais e garantir que a política seja voltada ao que realmente importa.

Na sequência, durante a formulação da política pública, entram em cena o uso de revisões sistemáticas sobre experiências anteriores e a elaboração da ferramenta Teoria da Mudança (TdM). A TdM, como veremos na Parte II deste guia, é uma ferramenta que organiza, de forma lógica e explícita, como se espera que os insumos e atividades da política levem à geração de produtos, resultados e impactos.

Com a TdM formulada, a Avaliação de Desenho permite uma análise crítica da concepção da política. Embora seja especialmente útil no momento ex ante, quando ainda há tempo de aprimorar o desenho antes da implementação, ela também pode ser realizada ex post, para investigar se eventuais falhas nos resultados decorreram de problemas estruturais no desenho. Em ambos os casos, o foco está na verificação da plausibilidade dos mecanismos causais, da fundamentação dos pressupostos e da coerência do desenho em relação ao contexto. Ao antecipar fragilidades e apontar ajustes necessários, essa avaliação contribui para tornar a política mais robusta e alinhada aos seus objetivos.

Esse é também o momento de realizar possíveis **simulações de cenários e projeções de impacto**, de planejar como será o **monitoramento da política** e de como ela será avaliada depois de implementada. O **plano de monitoramento e avaliação** precisa estar pronto **antes do lançamento da política**, devendo ser coerente com a TdM elaborada.

Análises ex ante têm, portanto, uma função estratégica: **ajudar a construir políticas mais consistentes, com maior potencial de impacto e melhor uso dos recursos públicos**. Elas garantem que o desenho da política tenha coerência interna, hipóteses plausíveis, metas claras e um caminho crível para a transformação desejada.

3.2 Análise ex post: compreendendo os efeitos da política após sua implementação

Depois que a política começa a ser implementada, entramos no campo das avaliações ex post. Nesse momento, a preocupação passa a ser entender como a política está sendo executadana prática e quais resultados está (ou não está) produzindo.Durante essa fase, utilizamos o **monitoramento contínuo** e realizamos **avaliações de implementação e de processo**, que permitem acompanhar a execução da política, detectar problemas operacionais, verificar se o público-alvo está sendo efetivamente alcançado e realizar ajustes em tempo real, caso necessário. Essas ferramentas apoiam a **gestão cotidiana da política**, favorecendo correções de rota e fortalecendo a capacidade de entrega da política. Mais adiante, já na etapa de **observação dos resultados e impactos**, entram as **avaliações ex post propriamente ditas**: avaliações de resultado, de impacto e de custo-benefício. Elas nos dizem se a política teve efeito, **para quem funcionou**, se valeu a pena do ponto de vista dos custos envolvidos e quais aprendizados podem ser extraídos para o futuro.

Essas avaliações são fundamentais para subsidiar decisões sobre a **continuidade, reformulação, expansão ou encerramento** da política, bem como para alimentar um ciclo virtuoso de aprendizado e melhoria contínua.

3.3 Análise ex ante ou ex post? Diferentes, complementares, essenciais

Esses dois momentos — ex ante e ex post — são distintos, mas **complementares**. Um não substitui o outro. Ao contrário: quando usados de forma articulada, oferecem uma visão completa da política pública, desde sua concepção até seus efeitos.

- A análise ex ante contribui para a **formulação da melhor política possível**, adequada ao problema e ao contexto, com coerência interna e potencial de impacto.
- A análise ex post permite **acompanhar política em ação**, corrigir rumos, avaliar impactos e prestar contas à sociedade.

Ambas fazem parte de uma gestão pública orientada por evidências, capaz de aprender com o passado e planejar melhor o futuro. Por isso, não existe um tipo de avaliação "melhor" em si. O que existe é o tipo mais apropriado à fase do ciclo em que a política pública se encontra — e à pergunta que se deseja responder.

A **Figura 3** apresenta essa lógica de forma visual, trazendo as principais perguntas avaliativas associadas a cada etapa do ciclo. A partir da pergunta, escolhemos a ferramenta de avaliação mais útil para o momento. Já a **Figura 4** sintetiza os tipos

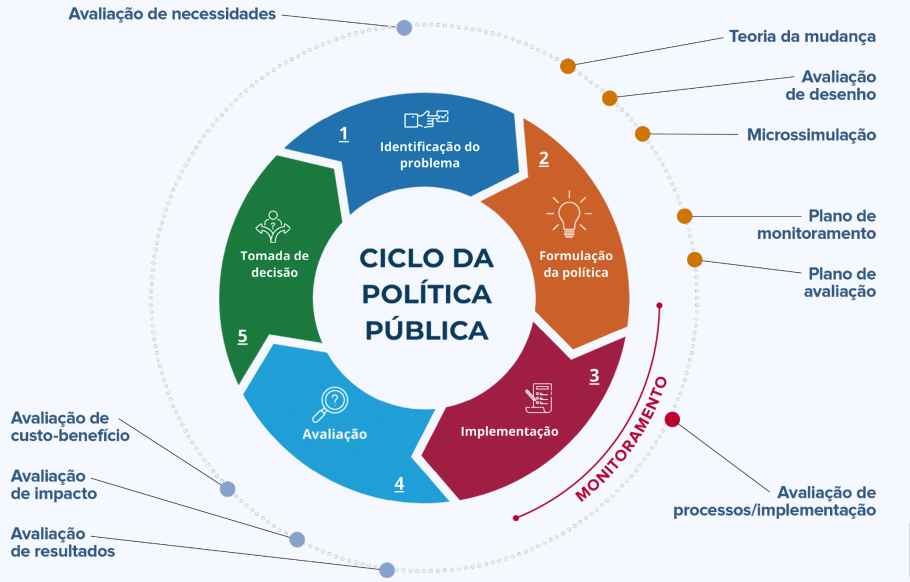
de avaliação mais indicados em cada fase, reforçando que a avaliação está presente ao longo de todo o ciclo — e não apenas ao final. A próxima seção detalha cada uma das etapas do ciclo, as perguntas avaliativas típicas de cada fase e os tipos de avaliação mais utilizados em cada uma delas.

Figura 3: O ciclo da política pública e exemplos de perguntas avaliativas em cada fase



Fonte: Elaborado com base em Secchi (2014) e Kraft e Furlong (2019).

Figura 4: Tipos de avaliação mais adequados em cada fase do ciclo da política pública.



Fonte: Elaborado com base em Secchi (2014) e Kraft e Furlong (2019).

3.4 Avaliação executiva: uma ferramenta transversal e ágil

Além das avaliações tradicionalmente associadas às etapas específicas do ciclo da política pública — como as avaliações de desenho, de implementação ou de impacto — existe um tipo de avaliação que se destaca por sua abrangência, agilidade e aplicabilidade transversal: a **avaliação executiva**.

Trata-se de uma metodologia estruturada que oferece um panorama estratégico e integrado da política pública, analisando múltiplas dimensões — como o diagnóstico do problema, a coerência do desenho, a qualidade da implementação e os efeitos observados — com base em dados existentes e métodos ágeis de coleta de dados. Por isso, embora seja classificada como um tipo de análise ex post, sua estrutura permite responder a perguntas avaliativas típicas de diferentes etapas do ciclo, sendo útil em momentos de revisão, reformulação, expansão ou tomada de decisão sobre a continuidade da política.

Por isso, é especialmente indicada quando há necessidade de gerar evidências de forma rápida, estruturada e com base em dados já disponíveis — apoiando decisões em momentos estratégicos da política. Essa modalidade é útil quando não há necessidade de compreender nenhuma dimensão de forma aprofundada, mas sim de obter um panorama geral que oriente decisões.

4 Etapas do ciclo e o uso de avaliações em cada fase

4.1 Identificação do problema

Essa etapa é o ponto de partida do ciclo. Buscamos compreender com profundidade o problema que deverá ser solucionado pela política: sua natureza, dimensão, grupos mais afetados, causas e consequências. Entre as perguntas centrais estão:

- Qual é o problema a ser enfrentado?
- Quais são os grupos mais afetados?
- Quais são suas causas e consequências?

Os tipos de avaliação mais adequados aqui são as **avaliações diagnósticas**, especialmente aquelas focadas em identificar as necessidades da população e as lacunas na resposta atual do poder público — também conhecidas como **avaliações de necessidades**.

4.2 Formulação da política

Uma vez compreendido o problema, é hora de desenhar a intervenção. Essa fase envolve as seguintes perguntas centrais:

- Quais são as **alternativas de políticas** que têm potencial para mitigar o problema enfrentado?
- Qual é a política mais adequada para o contexto?
- Qual é o **impacto esperado** da política?
- Qual é a lógica que embasa como a política pretende gerar os resultados esperados? Ou seja, qual é a **Teoria do Programa** e a **Teoria da Mudança**?
- Qual é o **plano de monitoramento e avaliação**?

Aqui, utilizamos evidências existentes para identificar alternativas de políticas. Para simulações de cenários e projeções de impacto, caso apropriado, existem as

metodologias de microssimulação. A partir da alternativa escolhida, empregamos a ferramenta da Teoria da Mudança para ilustrar a Teoria do Programa, desenhando a intervenção da forma mais robusta possível. A Teoria da Mudança, por sua vez, serve de base para a construção do plano de monitoramento e avaliação da política, que deve ser elaborado de forma coerente com essa estrutura.

4.3 Implementação

Nem sempre políticas bem formuladas são bem executadas. Por isso, é essencial acompanhar a entrega da política na prática. As perguntas típicas são:

- A política está sendo executada como planejado?
- As entregas foram de qualidade?
- O público-alvo está sendo atendido?
- Existem desvios ou falhas operacionais?

Nessa fase, utilizamos o **monitoramento contínuo** e as **avaliações de processo e implementação**, que ajudam a identificar ajustes necessários, corrigir rotas e apoiar a gestão. A avaliação de implementação verifica se o programa foi executado conforme o planejado, com foco na realização das atividades e no uso dos insumos. Já a avaliação de processo analisa como a política operou na prática, considerando a qualidade da entrega, a experiência dos usuários e os fatores contextuais que influenciaram a execução.

4.4 Avaliação de resultados, impactos e custos

Depois que a política foi implementada, é hora de entender seus efeitos concretos. Algumas perguntas centrais são:

- Os resultados esperados foram alcançados?
- A política causou um impacto desejado sobre os beneficiários?
- A política é custo-efetiva?

Essas respostas costumam ser obtidas por meio de avaliações de resultados, **avaliações de impacto e avaliações de custo-benefício e/ou custo-efetividade**. A avaliação de resultados, em geral, verifica se as metas de produtos e resultados foram alcançadas. A avaliação de impacto vai além: busca identificar se os efeitos almejados pela política foram, de fato, causados pela intervenção, isolando os efeitos da política de outros fatores. Já a avaliação de custo-benefício estima se

os benefícios gerados pela política superam os custos envolvidos, atribuindo valor monetário aos efeitos. Por fim, a avaliação de custo-efetividade compara os custos da política em relação aos efeitos alcançados sem necessidade de converter benefícios em valores monetários.

4.5 Tomada de decisão

Ao final do ciclo, as evidências geradas servem para informar decisões estratégicas sobre a continuidade, expansão, reformulação ou encerramento da política. Perguntas típicas incluem:

- A política deve continuar, ser ajustada ou encerrada?
- Como usar as evidências geradas para a tomada de decisão?

5 Redesenhando o ciclo: avaliação como eixo transversal

Tradicionalmente, a avaliação aparecer é representada como uma etapa final do ciclo da política pública, situada entre a implementação e a tomada de decisão. No entanto, essa visão é limitada: ela sugere que a avaliação só entra em cena no fim do processo, quando, na verdade, **avaliar é algo que deve acontecer desde o início** — e seguir acontecendo ao longo de toda a trajetória da política.

Por isso, propomos aqui um redesenho conceitual do ciclo, no qual a avaliação deixa de ser representada como uma fase isolada e passa a ser compreendida como um eixo transversal que acompanha todas as etapas da política. A etapa anteriormente intitulada "avaliação" passa a ser chamada de "resultados e impactos", refletindo com mais precisão o foco desse momento da política.

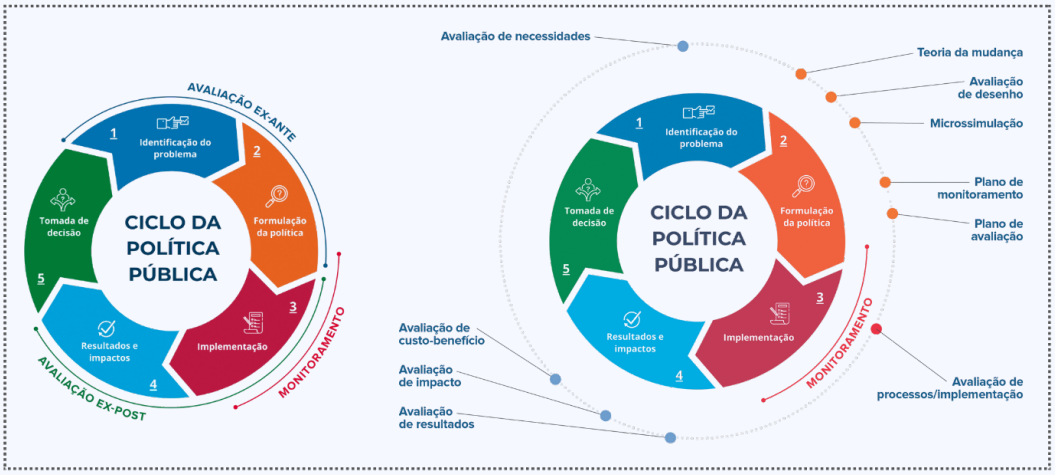
Com isso, o ciclo da política pública que orienta este guia é composto por cinco etapas principais:

1. Identificação do problema
2. Formulação da política
3. Implementação
4. Resultados e impactos
5. Tomada de decisão

Como demonstrado na **Figura 5**, a avaliação se mantém presente em todas as etapas do ciclo, ainda que de maneiras distintas, conforme os objetivos de cada fase. Seja para compreender melhor o problema, para desenhar soluções com base em evidências, para acompanhar a execução da política ou para medir seus efeitos, **avaliar é sempre um componente fundamental**.

Essa mudança de nomenclatura não é meramente estética. Ela traduz a ideia central deste guia: **a avaliação não é uma etapa pontual, mas uma prática contínua, adaptável e essencial à formulação e ao aprimoramento das políticas públicas**.

Figura 5: Avaliação ao longo de todo o ciclo da política pública



Fonte: FGV CLEAR, elaboração própria.

6 A abordagem do ciclo e a importância da institucionalização da avaliação nos governos de decisão

A adoção de uma abordagem integrada ao ciclo da política pública — com a avaliação presente desde o diagnóstico até a decisão final — **ajuda governos a estruturarem sistemas de avaliação para institucionalizar a gestão pública informada por evidências**. Trata-se de uma abordagem particularmente útil para governos que desejam qualificar decisões, aprimorar políticas e consolidar práticas duradouras de uso de evidências.

Ao organizarem seu processo decisório com base nas etapas do ciclo — identificando problemas, analisando alternativas, formulando políticas, monitorando a implementação e avaliando seus efeitos — os governos criam **condições institucionais para o uso sistemático de informações qualificadas**. Essa lógica estruturada favorece a transformação da avaliação de uma prática eventual em uma **função permanente e estratégica do Estado**.

Esse processo é o que chamamos de **institucionalização da avaliação**: a presença efetiva de estruturas, normas, competências e rotinas que permitem **produzir, utilizar e sustentar a prática avaliativa ao longo do tempo**, independentemente de governos específicos ou conjunturas políticas (FGV CLEAR, 2025).

A institucionalização da avaliação é fundamental para que a produção e o uso de evidências deixem de depender da iniciativa de indivíduos ou gestões pontuais, e passem a ser **parte integrante do funcionamento da administração pública**. Com isso, torna-se possível garantir que avaliações sejam feitas com regularidade, que seus achados sejam utilizados de forma prática na gestão e que a sociedade tenha acesso a informações claras sobre os efeitos das políticas públicas.

Box 1 – O que caracteriza a institucionalização da avaliação?

Governos que avançam na institucionalização da avaliação costumam apresentar cinco elementos-chave:

1. Legislação e marcos normativos:

   Existência de leis, decretos ou regulamentos que estabelecem a avaliação como dever da administração pública.

2. Estrutura institucional responsável:

   Órgãos, comitês ou unidades administrativas encarregadas de

coordenar e executar avaliações, preferencialmente próximos ao centro de governo.

3. Práticas avaliativas regulares:

   Avaliações planejadas e executadas de forma periódica e com critérios de qualidade definidos.

4. Capacidades técnicas e desenvolvimento de competências:

   Formação continuada de servidores, presença de equipes especializadas e produção de materiais técnicos.

5. Uso dos resultados para a gestão e o planejamento:

   Integração dos achados das avaliações à tomada de decisão, à alocação orçamentária e ao redesenho de políticas.

Fonte: FGV CLEAR (2025).

7 Conclusão

Incorporar a avaliação desde as etapas iniciais da política pública é uma forma de planejar de maneira estratégica e transparente, com foco nas reais necessidades da população. Quando o diagnóstico é bem conduzido, as alternativas são comparadas entre si com base nas evidências disponíveis, enquanto as decisões se apoiam em critérios claros e objetivos, o que aumenta significativamente as chances de a política alcançar os impactos desejados.

Ao integrar a avaliação desde o início, a política se torna mais coerente internamente, mais exequível na prática e mais suscetível a ajustes ao longo do tempo. Essa abordagem contribui para a construção de políticas públicas que aprendem, se adaptam e efetivamente melhoram a vida das pessoas.

Nos próximos capítulos, apresentamos com mais profundidade as ferramentas de avaliação que podem ser aplicadas ao longo do ciclo da política pública. São instrumentos técnicos, exemplos concretos e orientações práticas voltados a apoiar gestores e equipes na tarefa de planejar, monitorar, avaliar e aprimorar continuamente suas políticas — fortalecendo, assim, o uso estratégico de evidências na gestão pública.

PARTE I I

Análise *ex ante*: planejamento informado por evidências

Parte II – Análise ex ante: planejamento informado por evidências

1 Identificação e diagnóstico do problema público[5]

1.1 Por que é importante detalhar bem o problema?

Toda política pública nasce de um problema que precisa ser enfrentado. Por mais simples que essa ideia pareça, identificar e diagnosticar corretamente um problema público é uma das etapas mais complexas e, ao mesmo tempo, mais cruciais do ciclo da política pública. É a partir dessa etapa que todas as fases seguintes — a formulação, a implementação e a avaliação — serão estruturadas.

Definir o problema de forma clara, precisa e baseada em evidências ajuda a alinhar os esforços do governo, da sociedade civil e de outros atores envolvidos. Uma definição mal formulada pode levar a soluções inadequadas, desperdício de recursos e até ao agravamento da situação que se pretendia resolver. Por outro lado, um bom diagnóstico permite compreender quem é afetado, identificar as causas estruturais do problema, analisar seus efeitos sobre a sociedade e acompanhar sua evolução ao longo do tempo.

O uso qualificado de evidências torna esse processo mais robusto e objetivo. Com o apoio de dados, indicadores e escutas sistemáticas, é possível superar percepções vagas ou opiniões isoladas, oferecendo uma base sólida para a tomada de decisão. Além disso, o diagnóstico bem construído contribui para tornar a política pública mais transparente e compreensível para a sociedade, promovendo maior alinhamento entre governo e cidadãos.

Este capítulo apresenta um passo a passo prático para a identificação e o diagnóstico de problemas públicos, destacando como utilizar evidências para descrever o problema, identificar suas causas, compreender seus efeitos e definir sua relevância frente a outros desafios. Também são apresentadas ferramentas úteis — como o mapa de causas, a árvore do problema e as matrizes de priorização —, além de exemplos concretos que ilustram os conceitos ao longo do texto.

Nosso ponto de partida é simples, mas fundamental: não é possível desenhar uma política pública de forma adequada sem um problema bem definido.

[5] Este capítulo se baseia, de forma geral, na literatura técnica e experiências internacionais documentadas por Baltussen e Niessen (2006), BetterEvaluation (2024), Brasil (2018a), Colômbia (2021), Comissão Europeia (2024), IJSN (2018a), Kusek e Rist (2004), Neuman (2014), Patton (2008), PREA e SPC (2015), Veneklasen e Miller (2007), Watkins, Meiers e Visser (2012).

1.2 Etapas para a identificação e diagnóstico de um problema público

Diagnosticar bem um problema público exige método. Embora as etapas não precisem ser seguidas de forma totalmente linear, é importante contar com um roteiro estruturado que ajude a organizar o raciocínio, fundamentar o diagnóstico em evidências e orientar a tomada de decisão. O **Box 2** apresenta esse percurso, que forma a base do processo diagnóstico e será detalhado nas seções a seguir.

Box 2 – O que caracteriza a institucionalização da avaliação?

Delimitação do problema

↓

Levantamento de evidências

↓

Análise da distribuição do problema

↓

Avaliação da relevância e priorização

↓

Mapeamento de causas e consequências

Fonte: FGV CLEAR, elaboração própria.

1.2.1 Delimitação do problema

Definir claramente qual é o problema que deverá ser abordado pela política pública pode parecer trivial, mas esse passo exige rigor conceitual e atenção prática. Muitas vezes, o diagnóstico é apressado, e aquilo que se chama de "problema" já é, na verdade, uma proposta de solução — como "ampliar o acesso a creches" ou "construir novas escolas". Essas formulações já embutem juízos de valor e indicam caminhos de ação, sem antes esclarecer qual realidade concreta está sendo enfrentada.

Um problema público, nesse contexto, é uma situação observada, socialmente relevante e que demanda ação coletiva, geralmente por parte do Estado. Ele deve ser descrito de forma clara, objetiva e baseada em evidências, apontando:

- O tema em questão (por exemplo, frequência escolar, mortalidade infantil, crédito para pequenos negócios);
- Um ou mais indicadores que expressem sua magnitude;

- A meta ou estado desejado a ser alcançado.

Nessa etapa, o que deve orientar o processo não é ainda o "como agir", mas o "o que está acontecendo". Trata-se de investigar o fenômeno concreto que está na raiz do incômodo social, entendê-lo em sua complexidade e só então refletir sobre como enfrentá-lo. A pergunta central que deve guiar essa fase do diagnóstico é: qual realidade queremos compreender e transformar?

Para delimitar um problema público de maneira clara e fundamentada, é importante responder a um conjunto de perguntas orientadoras que ajudam a estruturar o diagnóstico com base em evidências. Essas perguntas funcionam como um passo a passo para garantir que o problema está bem descrito, quantificado, contextualizado e alinhado com os critérios de relevância social. O **Box 3** apresenta os principais elementos que devem estar presentes em uma boa delimitação do problema.

Box 3 – Elementos essenciais na delimitação de um problema

- O que está acontecendo? (Descrição objetiva)
- Desde quando? (Evolução temporal)
- Onde ocorre? (Abrangência geográfica)
- É quantificável? (Uso de indicadores)
- É socialmente relevante? (Impacto percebido)
- Qual a distância entre a situação atual e a desejada?

Fonte: FGV CLEAR, elaboração própria.

A primeira pergunta trata da descrição objetiva do que está acontecendo, ou seja, do fenômeno observado que gera incômodo ou prejuízo social. Em seguida, é preciso investigar desde quando essa situação ocorre, a fim de entender sua persistência e evolução ao longo do tempo. A análise da abrangência geográfica ajuda a dimensionar a escala do problema e identificar eventuais desigualdades territoriais. Verificar se o problema é quantificável, com o uso de indicadores, permite não apenas descrevê-lo com precisão, mas também acompanhar sua evolução. Avaliar sua relevância social garante que a política esteja alinhada às prioridades da população e aos valores coletivos. Por fim, estimar a distância entre a situação atual e a desejada ajuda a dimensionar o desafio e a orientar a definição de metas para a política pública. Responder a essas perguntas ajuda a construir um diagnóstico claro e consistente, conectando o problema à realidade concreta e ao contexto em que ele ocorre, conforme ilustrado no exemplo a seguir. A definição deve ser baseada em dados sempre que possível, permitindo mensurar a gravidade da situação e monitorar sua evolução ao longo do tempo.

Exemplo prático 1— Baixa cobertura de atendimento em creches

No município XYZ, 32% das crianças de 0 a 3 anos não frequentam creche ou qualquer forma de atendimento à primeira infância. Esse dado, registrado no diagnóstico de 2023, evidencia um quadro persistente de exclusão na oferta de serviços essenciais nessa etapa da vida. A meta do município, definida no Plano Plurianual, é reduzir esse número para abaixo de 15% até 2025.

Nesse caso, o problema público está claramente definido como: baixa cobertura de atendimento à primeira infância no município XYZ.
A formulação do problema explicita o indicador utilizado (proporção de crianças sem atendimento), o recorte territorial (município XYZ), a evolução temporal (com base no dado de 2023 e a meta de 2025), e a distância entre a situação atual e a desejada. O tema é socialmente relevante, pois o acesso à educação infantil é fundamental para o desenvolvimento das crianças e para a equidade de oportunidades.

Fonte: FGV CLEAR, elaboração própria.

Cuidado: problema não é solução

É comum encontrar definições de problema público que já indicam, de forma implícita, uma solução. Expressões como "falta de creches", "necessidade de aumentar a oferta de vagas" ou "construção de novas unidades" são exemplos frequentes desse tipo de formulação. Embora pareçam apontar para uma necessidade real, esses enunciados antecipam uma resposta antes mesmo de compreender plenamente a pergunta. Uma boa formulação do problema exige clareza, foco e o uso criterioso de dados. No entanto, para que essa definição não se baseie apenas em percepções, é fundamental sustentá-la com informações concretas e relevantes. Por isso, a próxima etapa do diagnóstico se dedica ao levantamento de dados e à análise de evidências — um passo essencial para transformar impressões em análises consistentes e orientar decisões mais bem informadas.

Aplicando na prática

- Defina o problema com base em dados e evidências, não apenas em percepções.
- Descreva o que está acontecendo, desde quando, e com quem.
- Evite termos vagos ou soluções embutidas como "falta de..." ou "necessidade de...".
- Use indicadores claros e, se possível, com metas e compa-

rações.

- Teste sua formulação com a pergunta: “Isso é um problema ou já é uma solução?”

1.2.2 Levantamento de evidências

Um diagnóstico bem-feito depende da qualidade das informações utilizadas para descrever o problema. Levantar, selecionar e interpretar adequadamente dados e indicadores é uma tarefa essencial para transformar percepções em análises fundamentadas.

Nesta etapa, o objetivo não é apenas reunir dados disponíveis, mas identificar **quais informações são úteis para entender o problema**, sua qualidade, onde estão localizadas e se podem ser utilizadas tal como estão ou exigem complementação. Ao sistematizar essas informações, cria-se a base para análises mais aprofundadas, tais como a identificação de grupos afetados, a localização do problema e suas possíveis causas. É por isso que o levantamento de dados **antecipa e alimenta todas as demais etapas do diagnóstico** e se conecta diretamente com as fases de formulação, implementação e avaliação da política pública.

As informações que sustentam o diagnóstico de um problema público podem ser organizadas em quatro grandes grupos, conforme explicitado no **Box 4**: fontes secundárias públicas, registros administrativos, levantamentos próprios e informações analíticas já produzidas.

As fontes secundárias incluem bases amplamente acessíveis, como aquelas disponibilizadas por órgãos oficiais e institutos de estatística. Já os registros administrativos são produzidos rotineiramente pela gestão pública e refletem a operação direta dos serviços, embora nem sempre estejam organizados para uso analítico.

Quando essas fontes são insuficientes, pode-se recorrer a levantamentos próprios, como pesquisas de campo ou entrevistas. Além disso, é importante considerar estudos já realizados — como avaliações, auditorias ou relatórios técnicos — que podem trazer aprendizados valiosos e evitar retrabalho. O uso combinado dessas diferentes fontes permite construir um diagnóstico mais robusto, completo e orientado à ação.

Box 4 – Fontes de informação úteis para o diagnóstico

As informações que apoiam o diagnóstico podem ser agrupadas em quatro tipos principais:

- Fontes secundárias públicas

  Bases de dados disponíveis para consulta aberta, organizadas

por órgãos oficiais ou instituições de pesquisa. Exemplos: IBGE, Censos, PNAD, SNIS, Inep, PISA, Ipea e Banco Mundial.

- Registros administrativos

  Dados produzidos no cotidiano da gestão pública, como prontuários, matrículas escolares, boletins de ocorrência e cadastros de programas. São fontes valiosas, mas nem sempre estão organizadas para uso imediato.

- Levantamentos próprios

  Quando os dados existentes são insuficientes ou inexistem, a equipe pode realizar coletas específicas, como entrevistas, grupos focais, visitas de campo ou questionários padronizados.

- Informações analíticas já produzidas

  Avaliações anteriores, auditorias, relatórios técnicos, pesquisas acadêmicas ou estudos de caso. Fontes como essas podem trazer aprendizados importantes sobre o problema e as respostas já tentadas.

Fonte: FGV CLEAR, elaboração própria.

O uso combinado de diferentes fontes de informação enriquece o diagnóstico. É útil considerar duas dimensões principais: primeiro, se os dados são estatisticamente representativos de uma população; segundo, se trazem informações sobre fatos objetivos (como frequência escolar, renda ou mortalidade) ou sobre percepções, narrativas e interpretações locais (como opiniões sobre a qualidade de um serviço ou barreiras sentidas pelas famílias). Enquanto os dados factuais permitem medir e comparar situações, os dados de natureza interpretativa ajudam a revelar dimensões menos visíveis e mais contextuais. Além disso, a incorporação de análises anteriores evita retrabalho e oferece insumos valiosos já sistematizados para orientar a reflexão e o aprofundamento do diagnóstico.

Exemplo prático 2— Descrevendo melhor a realidade com diferentes fontes

No município XYZ, a gestão identificava como problema a "baixa cobertura de atendimento à primeira infância". O único dado disponível era o número de crianças matriculadas nas creches públicas. Ao investigar melhor, percebeu-se que esse número não incluía creches conveniadas nem instituições comunitárias apoiadas pelo poder público. Também não havia estimativa precisa da população de crianças de 0 a 3 anos por bairro. A partir do cruzamento entre registros

administrativos da educação, dados do Cadastro Único e estimativas populacionais do IBGE, foi possível construir um indicador mais realista de cobertura — e identificar os territórios com maior déficit. A equipe complementou essa análise com entrevistas realizadas com as mães em fila de espera e lideranças comunitárias, revelando barreiras de acesso, distância e falta de informação que os dados sozinhos não captavam.

Fonte: FGV CLEAR, elaboração própria.

O exemplo acima ilustra como o levantamento de informações vai muito além da coleta de dados: ele exige olhar crítico, articulação entre fontes diversas e sensibilidade para captar aspectos que nem sempre aparecem nos números. Um diagnóstico robusto combina informações já disponíveis com análises específicas da realidade local, valorizando tanto os dados quantitativos quanto as evidências qualitativas, o que ajuda a contextualizar e interpretar os padrões observados. A seguir, reunimos algumas orientações práticas para apoiar gestores e equipes na condução dessa etapa fundamental do processo diagnóstico.

Usar boas evidências não significa apenas reunir dados, mas selecionar criteriosamente informações confiáveis, relevantes e bem definidas. Quando um indicador é construído com base em dados de baixa qualidade — incompletos, desatualizados ou imprecisos , ele acaba sendo uma representação distorcida da realidade. E, se partimos de uma representação equivocada, corremos o risco de traçar diagnósticos errados e propor soluções ineficazes ou até contraproducentes. Por isso, é fundamental avaliar cuidadosamente cada fonte: sua origem, frequência de atualização, método de cálculo e cobertura. A qualidade da informação é tão importante quanto sua existência — e deve ser tratada como parte central do processo analítico.

Aplicando na prática

- Comece pelas fontes já existentes: bases públicas, registros internos, estudos e avaliações anteriores.
- Avalie a qualidade, atualidade e relevância dos dados antes de utilizá-los.
- Prefira indicadores desagregáveis por sexo, idade, raça/-cor e território.
- Use dados qualitativos para complementar e interpretar os quantitativos.
- Registre as fontes e limitações de cada informação usada.

- Planeje bem qualquer nova coleta de dados para evitar desperdícios de tempo e recursos.

1.2.3 Análise da distribuição do problema

Quem é mais afetado?

Uma vez delimitado o problema, o passo seguinte é compreender **quem são os grupos sociais e territórios mais impactados por ele**. Essa análise é essencial para orientar o foco da política pública, priorizar recursos e garantir que a resposta da política esteja à altura da urgência e da localização do problema.

A manifestação de um problema raramente é homogênea. Mesmo quando atinge uma parcela ampla da população, seus efeitos costumam ser mais severos para certos grupos — como pessoas em situação de pobreza, mulheres, população negra ou indígena, crianças, idosos ou moradores de determinadas regiões. **Fatores como renda, gênero, raça/cor, faixa etária e território** influenciam tanto a exposição ao problema quanto a capacidade de enfrentá-lo.

Essa identificação começa, em geral, com a **desagregação dos dados disponíveis**. Bases administrativas, cadastros sociais e pesquisas amostrais costumam trazer informações que permitem examinar o problema por faixa etária, sexo, cor/raça, território e condição socioeconômica. Essa análise revela desigualdades que não aparecem quando são examinadas apenas as médias agregadas.

Exemplo prático 3— Médias que escondem desigualdades

Uma rede municipal de ensino comemorava um desempenho "acima da média" em avaliações de proficiência em leitura e matemática. No entanto, ao desagregar os dados por território, identificou-se que escolas em bairros periféricos apresentavam resultados drasticamente inferiores. Em uma das regiões, menos de 30% dos alunos do 5º ano atingiam o nível básico de proficiência em matemática, enquanto a média municipal era de 65%. Esse dado foi fundamental para redirecionar os esforços da política educacional para os territórios da periferia da cidade.

Fonte: FGV CLEAR, elaboração própria.

Além da análise de dados e indicadores quantitativos, é importante ouvir as pessoas que convivem com o problema. Métodos qualitativos, como **entrevistas, grupos focais e oficinas participativas** ajudam a revelar dimensões subjetivas, barreiras práticas e percepções sociais que não estão nas estatísticas. Conversar com os grupos afetados permite enxergar o problema a partir de diferentes perspectivas e fortalecer a legitimidade do diagnóstico.

Aplicando na prática

- Desagregue os dados por sexo, raça/cor, idade, território e renda.
- Avalie se os efeitos do problema são mais intensos para certos grupos.
- Não se baseie apenas em médias: procure desigualdades entre grupos e territórios.
- Combine dados quantitativos com escuta qualificada de atores relevantes.
- Verifique se novas coletas de dados são realmente necessárias e planeje cuidadosamente, para evitar desperdícios de tempo e recursos.

Comparar para compreender: o que acontece em outros lugares?

Além de identificar quem é mais afetado, é útil compreender onde o problema se concentra e como ele se manifesta em relação a outros territórios ou metas de referência. A comparação territorial e temporal fortalece o diagnóstico porque permite:

- Evidenciar desigualdades espaciais dentro do território analisado;
- Entender a gravidade do problema em relação a outros contextos, como municípios vizinhos, médias estaduais ou nacionais;
- Dimensionar a distância até metas de longo prazo, como os Objetivos de Desenvolvimento Sustentável (ODS).

Exemplo prático 4— Comparação que muda o diagnóstico e dimensiona o desafio

Uma secretaria municipal de educação avaliava como preocupantes os resultados dos alunos em leitura e matemática, com apenas metade dos estudantes atingindo o nível básico de proficiência nas avaliações padronizadas do Sistema de Avaliação da Educação Básica (SAEB). A situação parecia grave. No entanto, ao comparar com a média nacional, identificou-se que a proficiência em leitura estava acima da média brasileira e em níveis satisfatórios— enquanto os resultados em matemática estavam significativamente abaixo. Essa comparação permitiu reformular o diagnóstico: o problema não era "baixo desempenho em geral", mas um desafio específico e mais crítico na área de matemática. A equipe técnica também utilizou como referência a **Meta 4.6**

dos Objetivos de Desenvolvimento Sustentável (ODS) — que estabelece como objetivo garantir que **todos os jovens – e uma proporção substancial de adultos, homens e mulheres – alcancem proficiência em leitura, escrita e aritmética** — como referência para estimar a distância entre a situação atual e o patamar desejado. Isso ajudou a dimensionar o esforço necessário, dando mais clareza e legitimidade ao diagnóstico e à definição de prioridades para a política educacional.

Fonte: FGV CLEAR, elaboração própria.

A comparação, portanto, não serve apenas para posicionar o problema, mas para **enriquecer o diagnóstico com perspectiva contextual**. Ao analisar como o problema aparece em diferentes locais ou frente a metas reconhecidas, ampliamos nossa capacidade de interpretar sua gravidade, urgência e especificidades. Essa etapa do diagnóstico contribui para decisões mais realistas, fundamentadas e coerentes com a realidade local — e prepara o caminho para pensar em estratégias mais efetivas de enfrentamento.

Aplicando na prática

- Compare dados entre bairros, municípios, regiões ou países.
- Observe como o problema varia em relação a médias estaduais ou nacionais.
- Use metas nacionais ou globais (como os ODS) para entender a distância entre a situação atual e a desejada.
- Use a comparação para orientar prioridades e calibrar expectativas.

1.2.4 Avaliação da relevância do problema e priorização

Na prática da gestão pública, os problemas raramente aparecem de forma isolada. É muito comum que equipes técnicas se deparem com múltiplos problemas legítimos, todos com impactos relevantes e evidências que justificam atenção. No entanto, recursos financeiros, capacidade institucional e tempo político são limitados. Por isso, definir quais problemas serão priorizados, por onde começar e o que poderá ser endereçado em um segundo momento é uma decisão estratégica — e inevitável.Essa decisão envolve análise técnica e julgamento político. Problemas mais graves nem sempre são aqueles com maior viabilidade de enfrentamento. Da mesma forma, problemas que geram forte demanda social nem sempre estão bem caracterizados tecnicamente. O papel do gestor, nesse cenário, é conciliar esses

diferentes elementos com responsabilidade, clareza de critérios e transparência das escolhas feitas.

- **Priorizar um problema é escolher com base em critérios**

Para evitar decisões baseadas apenas em percepções subjetivas ou pressões momentâneas, o processo de priorização deve considerar múltiplas dimensões. Esses elementos podem ser organizados em **critérios técnicos e estratégicos**, como:

  - Gravidade do problema: número de pessoas afetadas, severidade dos efeitos, efeitos em cadeia.
  - Urgência da intervenção: risco de agravamento, janelas de oportunidade, perdas irreversíveis.
  - Viabilidade de ação: existência de instrumentos legais, recursos disponíveis, competências institucionais.
  - Pressão social e legitimidade: presença de demandas da população, mobilização de partes interessadas, atenção política.
  - Sinergia com outras políticas: capacidade de a solução desencadear efeitos positivos sobre outros problemas.
  - Alinhamento com planos e compromissos oficiais: inclusão em metas, pactos ou estratégias já aprovadas.

Esses critérios podem ser discutidos em equipe ou com apoio de instâncias intersetoriais, como conselhos, grupos técnicos ou espaços participativos. O uso de matrizes de priorização, que atribuem pesos a cada critério, pode ser uma forma útil de estruturar a decisão sem engessá-la (VENEKLASEN & MILLER, 2007).

Exemplo prático 5— Quais problemas da educação vêm primeiro?

Em um município de médio porte, a equipe da Secretaria de Educação identificou, com base em dados e escutas, cinco desafios relevantes:

- Baixa frequência escolar no ensino fundamental
- Baixo desempenho em matemática no ensino fundamental
- Déficit de vagas em creches
- Infraestrutura escolar deteriorada em parte da rede
- Ausência de formação continuada para professores

Todos os desafios são reconhecidos como importantes — tanto pelos dados quanto pelas percepções da comunidade escolar. Mas o orçamento anual só permite investimentos significativos em duas frentes.

A equipe realizou uma oficina de priorização com técnicos e gestores. Após aplicar alguns dos critérios discutidos — gravidade, urgência, viabilidade, legitimidade e sinergia —, a decisão foi priorizar:

- A baixa frequência escolar no ensino fundamental: por ser um problema com consequências imediatas sobre o aprendizado, está fortemente associada à desigualdade social e ao risco de evasão. Os dados mostraram que a frequência escolar era especialmente baixa entre alunos de famílias em situação de vulnerabilidade. A equipe considerou que havia evidências sólidas sobre intervenções eficazes — como oferta de merenda escolar, transferência de renda condicionada e uso de busca ativa — e capacidade local para implementar tais estratégias com agilidade.
- A expansão da educação infantil, para enfrentar o déficit de vagas em creches, por ser um problema que afeta diretamente o desenvolvimento na primeira infância e a inserção das mulheres no mercado de trabalho. Havia projeto executivo pronto e recursos federais disponíveis, o que aumentava a viabilidade imediata da intervenção.

As demais frentes — desempenho em matemática, infraestrutura escolar e formação continuada — foram incluídas no planejamento estratégico da secretaria para os anos seguintes, com ações preparatórias em andamento.

Essa escolha mostra que priorizar é alinhar capacidade, impacto e estratégia, sem abandonar os demais problemas. Ao selecionar frentes com maior potencial transformador e viabilidade de execução, a política pública ganha foco, efetividade e chances reais de entregar resultados positivos à população. O mais importante é criar um processo transparente no qual as escolhas são criteriosas e justificadas.

Fonte: FGV CLEAR, elaboração própria.

Como usar uma matriz de priorização?

Uma ferramenta bastante útil para apoiar esse tipo de escolha é a matriz de priorização. Trata-se de uma tabela simples em que os problemas são organizados em linhas, e os critérios de decisão em colunas. A equipe atribui notas (por exemplo, de 1 a 5) para cada problema, considerando critérios como, por exemplo, gravidade, urgência, viabilidade de ação, visibilidade pública ou pressão social e impacto sobre grupos vulneráveis.

Ao final, a pontuação total ajuda a visualizar quais problemas reúnem maior peso coletivo. Não se trata de uma fórmula matemática, mas de um instrumento que estrutura o debate, reduz o risco de decisões arbitrárias e favorece o alinhamento entre áreas técnicas e a direção política da política pública.

A matriz pode ser usada em reuniões técnicas, oficinas de planejamento ou em processos participativos mais amplos — especialmente úteis em contextos com múltiplos atores envolvidos na decisão. Os critérios e pesos podem ser ajustados conforme o contexto. A pontuação total é um indicativo de prioridade, mas deve ser complementada com análise estratégica, capacidade de ação e articulação institucional (VENEKLASEN & MILLER, 2007; WATKINS, MEIERS & VISSER, 2012).

Exemplo prático 6— Matriz de priorização

Problema identificado	Gravidade (1–5)	Urgência (1–5)	Viabilidade (1–5)	Impacto sobre vulneráveis	Pressão social (1–5)	Total de pontos
Déficit de vagas em creches	5	4	5	5	4	23
Baixo desempenho em matemática	4	3	4	5	3	19
Infraestrutura escolar precária	3	3	2	5	4	17

Fonte: Elaboração própria, adaptado a partir de VeneKlasen e Miller (2007) e Watkins, Meiers & Visser (2012).

Aplicando na prática

- Liste todos os problemas identificados e seus impactos, com base em dados e consultas com partes interessadas;
- Use critérios técnicos e políticos para comparar e justificar a escolha;
- Leve em conta a capacidade de agir, o tempo disponível e o contexto institucional;
- Envolva diferentes atores no processo para fortalecer a legitimidade da decisão;
- Registre os motivos da priorização e planeje estratégias futuras para os demais problemas.

1.2.5 Mapeamento de causas e consequências: a árvore do problema como ferramenta de diagnóstico

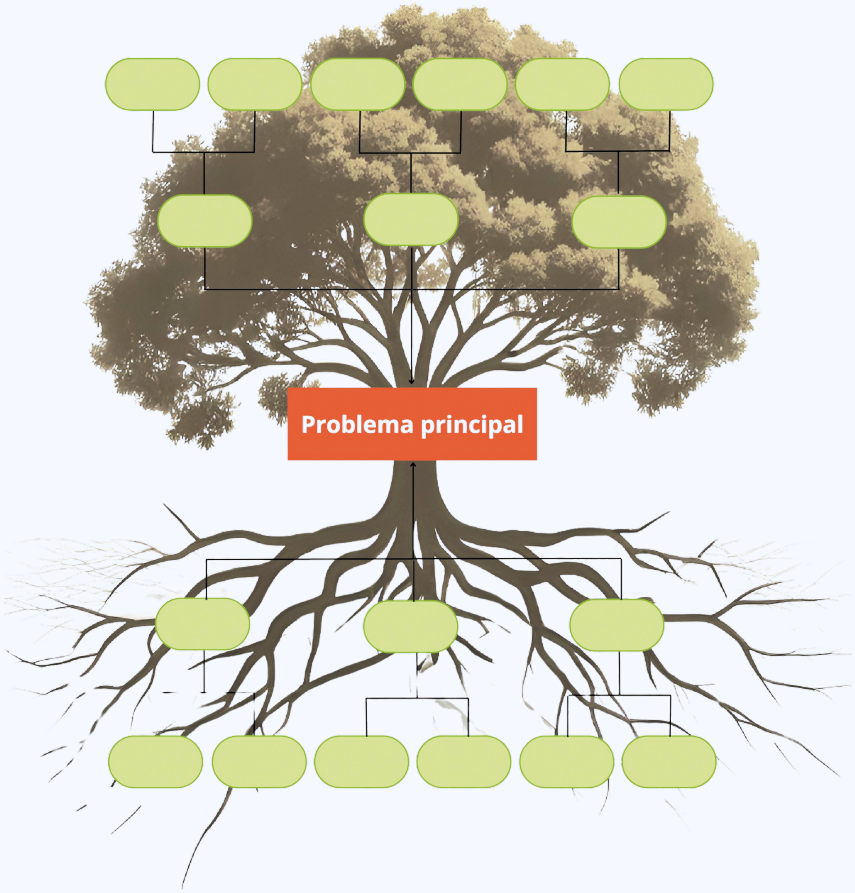
Uma vez definido qual problema será priorizado, é hora de aprofundar o diagnóstico. A escolha de um problema não encerra o processo de análise — ao contrário, marca o início de uma nova etapa: compreender por que ele ocorre, como se sustenta ao longo do tempo e quais efeitos provoca sobre a sociedade. Para isso, é essencial mapear suas causas e consequências de forma estruturada. A ferramenta mais utilizada nesse momento é a árvore do problema, também chamada de diagrama de causas e efeitos ou mapa de causas.

A árvore do problema é uma representação visual que organiza as relações entre **causas**, o **problema central** e suas **consequências**. Nessa estrutura, o tronco representa o problema principal; as raízes, suas causas; e os galhos, suas consequências. Essa organização favorece a construção de um diagnóstico compartilhado entre gestores e equipes técnicas, promovendo uma compreensão articulada dos elementos envolvidos. Ao contrário de uma lista solta de fatores, a árvore explicita **relações causais**, o que permite formular soluções mais coerentes.

A metáfora da árvore nos ajuda a **visualizar as dinâmicas que sustentam e propagam um problema público**. Assim como uma árvore real depende de raízes para se manter de pé, **os problemas sociais persistem porque têm origens profundas e interligadas**. Essas causas nem sempre são visíveis à primeira vista, mas alimentam o problema central. Por outro lado, ele também gera efeitos — ou seja, consequências que se espalham, como galhos, e afetam outros aspectos da vida social. Ao organizar essas relações em um diagrama, conseguimos compreender melhor como o problema se forma, se sustenta e se propaga.

Essa representação evita que a política pública se concentre apenas em resolver as consequências — como a evasão escolar, a superlotação hospitalar ou o desemprego — **sem atacar o que realmente causa esses problemas**. Soluções que tratam apenas as consequências visíveis são paliativas: podem gerar alívio momentâneo, mas não resolvem a origem da questão.

Figura 6: Árvore do problema



Fonte: FGV CLEAR, elaboração própria.

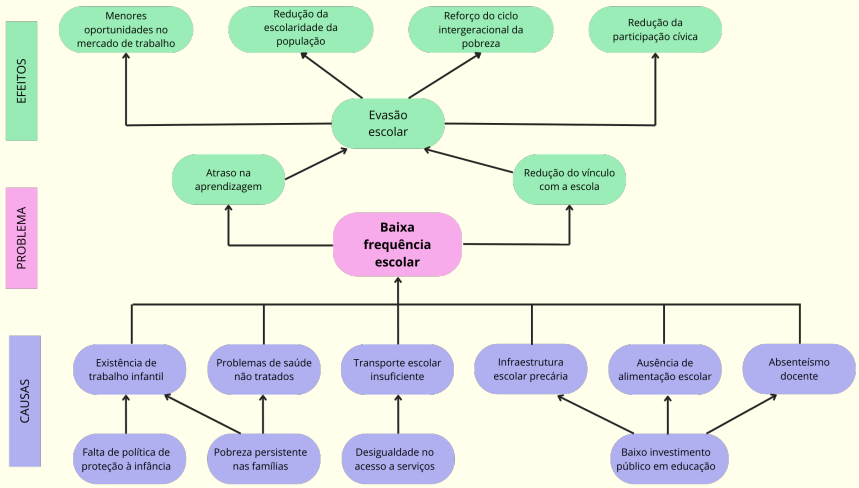
Exemplo prático 7— A árvore do problema da baixa frequência escolar

A **Figura 7** ilustra uma árvore do problema construída para diagnosticar a baixa frequência escolar no ensino fundamental. O tronco do diagrama representa o problema central. Abaixo dele, as raízes organizam as causas diretas, — como a ausência de alimentação escolar, o transporte insuficiente, as condições precárias de infraestrutura e o trabalho infantil. Essas causas estão conectadas a fatores mais profundos, como a pobreza das famílias, a desigualdade territorial e o baixo investimento em educação básica.

Na parte superior da árvore, aparecem estão os efeitos:s do problema: atraso na aprendizagem e evasão escolar aparecem como efeitossão consequências imediatas, que, por sua vez, . imediatos. Esses, por sua vez, levam a consequências impactos de longo prazo, como menores oportunidades no mercado de trabalho, perpetuação do ciclo da pobreza e redução da escolaridade média da população.

Esse tipo de estrutura ajuda a equipe a perceber que, se a política se limitar, por exemplo, à reforma da infraestrutura escolar ou à compra de computadores, dificilmente o quadro geral se transformará. A baixa frequência escolar não é um desafio restrito ao ambiente educacional, mas está enraizada em por fatores sociais, econômicos e institucionais que exigem intervenções voltadas às causas prioritárias.

Figura 7 – Árvore do problema da baixa frequência escolar



Fonte: FGV CLEAR, elaboração própria.

Box 5 – Por que usar a árvore do problema?

- Ajuda a decompor problemas complexos em causas específicas e interligadas;
- Explicita relações de causa e efeito que nem sempre são visíveis;
- Evita soluções que tratam apenas as consequências, e não as origens;
- Facilita o engajamento de atores diversos na construção do diagnóstico;
- Conecta o diagnóstico à formulação de objetivos, indicadores e estratégias.

Fonte: FGV CLEAR, elaboração própria.

Como identificar causas prioritárias?

Ao analisar a árvore construída, surge uma pergunta inevitável: **por onde começar?** Nem sempre será possível intervir em todas as causas ao mesmo tempo. Por isso, é preciso estabelecer prioridades — ou seja, definir quais causas são prioritárias e merecem atenção imediata.

Para fazer essa priorização, é importante definir alguns critérios que orientem a escolha de forma estruturada e transparente. A literatura sugere diversas possibilidades, que podem ser adaptadas ao contexto de cada política (BALTUSSEN & NIESSEN, 2006; WATKINS, MEIERS & VISSER, 2012). Alguns exemplos são:

- **Magnitude do impacto**: quanto a causa contribui para o problema central;
- **Capacidade de intervenção**: existência de instrumentos, autonomia institucional e recursos para enfrentamento;
- **Evidência disponível**: dados, avaliações ou estudos que indiquem que intervir nessa causa tende a gerar resultados concretos;
- **Urgência**: risco de agravamento rápido ou presença de janela de oportunidade;
- **Sinergia**: possibilidade de resolver outras causas ou mitigar efeitos ao atacar essa causa;
- **Cobertura por políticas existentes**: existência de ações em curso que já atacam essa causa, direta ou indiretamente;
- **Alçada institucional**: qual órgão é responsável pela causa e se há capacidade de coordenação intersetorial.

Nem sempre a causa mais importante será a mais viável para ação imediata. E nem sempre a mais viável será a que produz maior impacto. O exercício de priorização exige **análise técnica combinada com decisão estratégica**, e pode ser apoiado por ferramentas como matrizes de pontuação, oficinas intersetoriais ou simulações de cenários.

Exemplo prático 8— Como priorizar causas: o caso da mortalidade no trânsito

Em um município com altos índices de mortes no trânsito, a equipe técnica elaborou uma árvore do problema para entender melhor a situação. O problema central era a elevada letalidade de acidentes. As causas identificadas incluíam: imprudência dos motoristas (como excesso de velocidade e embriaguez), sinalização deficiente, fiscalização irregular, limites de velocidade elevados nas vias urbanas e ausência de campanhas educativas.

A partir do mapeamento, a gestão precisava decidir por onde começar a intervir. Nem todas as causas podiam ser enfrentadas ao mesmo

tempo. A equipe aplicou três critérios de priorização:

1. Efeito sobre o problema: qual causa contribui mais diretamente para as mortes?
2. Capacidade de intervenção: o município tem autonomia e instrumentos para agir sobre essa causa?
3. Evidência de efetividade: há dados ou experiências que mostram que intervir nessa causa funciona?

A imprudência ao volante, especialmente por embriaguez e excesso de velocidade, foi identificada como a causa mais crítica. Recorrente nos dados, essa conduta poderia ser enfrentada por meio de fiscalização e campanhas educativas, com evidências indicando a eficácia dessas estratégias. Por isso, foi escolhida como prioritária. Já a proposta de reduzir os limites de velocidade em determinadas vias, embora justificada, exigiria alterações legislativas e enfrentaria forte resistência popular. Por isso, a gestão avaliou que não havia, naquele momento, viabilidade política para seguir com essa mudança. A causa não foi descartada, mas foi postergada como parte de uma estratégia de médio prazo.

Esse processo mostrou como a árvore do problema pode ajudar a alinhar o diagnóstico à capacidade real de ação, permitindo decisões mais realistas e impactantes.

Fonte: FGV CLEAR, elaboração própria.

Ao final, a árvore do problema deve permitir que a equipe gestora responda com clareza: **O que está causando o quê? Onde faz mais sentido intervir primeiro? E com quais instrumentos?**

Essa representação ajuda a transformar o diagnóstico em estratégia, permitindo que a política pública se concentre **nas causas certas, no momento certo**. Mais do que listar fatores, a árvore orienta escolhas: ela permite **atacar as raízes — e não apenas os ramos** do problema, alocando recursos nas frentes com maior potencial transformador e viabilidade de implementação.

Aplicando na prática

- Escolha um problema central bem definido, claro e mensurável;
- Use a pergunta “por que isso acontece?” para identificar causas;

- Use a pergunta “o que isso provoca?” para identificar consequências;
- Organize as informações em uma árvore composta por três partes: raízes, tronco e galhos;
- Valide a árvore com atores diversos, como gestores, técnicos e representantes da população;
- Use a árvore como base para definir prioridades, objetivos e estratégias da política pública.

2 Formulação da Política: escolhas estratégicas para enfrentar o problema[6]

A formulação de uma política pública é o momento de definir como a intervenção vai funcionar: quais objetivos serão perseguidos, quais ações serão executadas e quais resultados se espera alcançar. Por isso, é comum que equipes técnicas e gestores queiram começar por aqui.

Em muitos casos, o desenho da política é guiado por intenções legítimas, mas não está suficientemente ancorado em evidências sobre o problema e seu contexto. Isso compromete a efetividade da intervenção e limita seus resultados. Antes de avançar para a formulação, é fundamental ter clareza sobre qual problema será enfrentado, suas causas e suas repercussões. Só depois de realizar esse percurso — como mostramos no Capítulo 1 — é possível construir uma solução com maior potencial de impacto positivo.

A partir de agora, este capítulo parte do pressuposto de que o problema já foi bem delimitado, analisado e priorizado. A seguir, focamos em como transformar esse diagnóstico em ação.

2.1 Buscando alternativas de políticas com base em evidências

Uma vez que o problema foi diagnosticado com clareza, o próximo passo é pensar em quais caminhos podem ser adotados para enfrentá-lo. Essa etapa da formulação exige criatividade, mas também realismo. Antes de elaborar algo novo, é essencial entender o que já foi tentado — no seu território, em outros estados, em países semelhantes — e quais os resultados alcançados. O objetivo é ampliar o repertório da equipe, aprender com experiências anteriores e evitar erros já cometidos por outros.

A formulação de alternativas deve idealmente partir da identificação de diferentes soluções previamente implementadas e avaliadas. Basear-se apenas na intuição ou em exemplos isolados pode levar a escolhas inadequadas ou pouco efetivas. É preciso fundamentar a decisão em evidências robustas e adequadas ao contexto. A política pública é sempre implementada em um contexto específico — por isso, o que funcionou em um lugar pode não ter o mesmo efeito em outro. Avaliar alternativas exige atenção tanto à efetividade demonstrada quanto à aderência ao contexto local.

[6]Este capítulo se baseia nos trabalhos de Al Noman, et al. (2024), Brasil (2018a), IJSN (2018a), Imas e Rist (2009), Kraft e Furlong (2019), OCDE (2023), Reino Unido (2022), TCU (2020).

2.1.1 Ponto de Partida: Delimitando o público da política: população potencial, objetivo e beneficiária

Definir quem será atendido por uma política pública é uma das tarefas centrais da etapa de formulação. Esse processo exige cuidado técnico, pois envolve escolhas que afetam diretamente a efetividade, a equidade e a viabilidade da política. Como toda política opera sob restrições de recursos, não é possível atender simultaneamente todas as pessoas afetadas por um problema público. Por isso, é necessário estabelecer critérios claros de **focalização** e **priorização**.

Esses dois conceitos, embora relacionados, não são sinônimos:

- **Focalização** diz respeito à definição de quem tem direito ou se encaixa nos critérios para ser atendido — ou seja, quem é elegível. Essa é a base para definir a **população objetiv**o, o grupo que a política busca afetar diretamente com sua intervenção.
- **Priorização** refere-se à forma de organizar o atendimento dentro da população objetivo, quando os recursos ou a capacidade de implementação não permitem atingir a todos de imediato. Isso define a **população beneficiária**, ou seja, quem será efetivamente atendido em um determinado momento ou fase da política.

Em outras palavras:

- A **população potencial** inclui todas as pessoas afetadas direta ou indiretamente pelo problema.
- A **população objetivo** é o subconjunto da população potencial que se deseja alcançar com a política, com base em critérios de elegibilidade. Ela representa o público-alvo da ação.
- A **população beneficiária** é a parcela da população objetivo que será de fato atendida, conforme critérios de priorização como ordem de inscrição, sorteio, filas ou distribuição geográfica.

A população objetivo é, portanto, o grupo que se pretende beneficiar ao longo do tempo. Ela pode ainda não ser atendida no momento inicial, mas foi definida como público-alvo da política. Já a população beneficiária representa quem está sendo efetivamente impactado em cada fase ou ciclo da implementação.

A definição desses grupos deve se apoiar em critérios objetivos e verificáveis — como região, faixa etária, situação de vulnerabilidade, nível de renda ou condição de saúde — para garantir transparência, reduzir ambiguidades e facilitar a identificação dos beneficiários.

Além de definir quem será atendido, é fundamental que os formuladores documentem com clareza os critérios adotados para a focalização e a priorização. As escolhas feitas devem ser devidamente justificadas, com base em evidências, e alinhadas aos objetivos da política. Também é importante estimar o tamanho e o perfil socioeconômico dos grupos definidos — população potencial, objetivo e beneficiária. Essas informações são cruciais para embasar o desenho da política, orientar a escolha entre alternativas disponíveis e permitir um planejamento mais preciso da implementação.

2.1.2 Onde encontrar boas evidências?

Existem várias formas de buscar evidências concretas para inspirar o desenho de uma política pública. O ideal é buscar em diferentes fontes que documentem experiências anteriores e ajudem a compreender o que funcionou (ou não) em contextos semelhantes.

Há hoje um conjunto crescente de bases de dados e repositórios que organizam evidências de políticas públicas testadas com métodos rigorosos. Muitos deles oferecem **resumos acessíveis dos achados mais relevantes**, organizados sob a forma de policy briefs, revisões sistemáticas e meta-análises — ferramentas úteis para apoiar decisões rápidas e bem informadas.

- **Policy briefs** são documentos curtos, produzidos para tomadores de decisão, que sintetizam as principais conclusões de uma pesquisa ou grupo de estudos sobre determinado tema.
- **Revisões sistemáticas** reúnem e analisam criticamente todos os estudos existentes sobre uma determinada intervenção, seguindo critérios metodológicos padronizados.
- **Meta-análises** vão além: quantificam os efeitos médios observados nos estudos analisados, permitindo comparações mais objetivas entre políticas.

Esses produtos são especialmente úteis para quem precisa de respostas práticas sobre o que funciona para alcançar o impacto pretendido em uma área específica. A seguir, no **Box 6**, listamos algumas das principais plataformas nacionais e internacionais que organizam essas evidências de forma clara, confiável e acessível.

Além disso, estudos acadêmicos, relatórios técnicos e avaliações publicadas por universidades, centros de pesquisa, organismos multilaterais ou think tanks especializados também podem oferecer insumos relevantes.

No próprio território, vale consultar programas já implementados, mesmo que não formalmente avaliados. Gestores e técnicos locais costumam ter percepções relevantes (e possivelmente divergentes) sobre o que foi feito, os resultados alcançados e os desafios enfrentados. No entanto, é importante lembrar que essas opiniões podem refletir visões subjetivas ou casos isolados e, por isso, devem ser consideradas com cautela.

Outra estratégia é investigar políticas semelhantes adotadas em outros municípios, estados ou países com contextos comparáveis. Experiências documentadas ou avaliadas em diferentes territórios ampliam o repertório da equipe e ajudam a identificar riscos e boas práticas. Especialistas da área podem ser aliados valiosos nesse processo. Em geral, estão familiarizados com o conjunto de evidências teóricas e práticas disponíveis e podem indicar estudos relevantes, resultados pouco divulgados ou aspectos contextuais que merecem atenção.

Box 6 – Onde encontrar boas evidências?

Essas plataformas reúnem estudos avaliativos, revisões sistemáticas e sínteses acessíveis sobre evidências do que funciona em políticas públicas.

No Brasil (em português)

- FGV CLEAR

  Centro regional da Global Evaluation Initiative que publica conteúdos sobre avaliação e boas práticas em políticas públicas nos países lusófonos.

  fgvclear.org

- Impacto IMDS

  Plataforma brasileira que reúne estudos avaliativos de políticas sociais com linguagem acessível.

  impacto.imdsbrasil.org

Internacionais

- J-PAL – Poverty Action Lab

  Rede global de pesquisadores que realiza e divulga avaliações de impacto em países em desenvolvimento.

  povertyactionlab.org

- IPA – Innovations for Poverty Action

  Conduz experimentos para testar soluções contra a pobreza em escala global.

  poverty-action.org

- 3ie – International Initiative for Impact Evaluation

  Produz revisões sistemáticas e bancos de dados com centenas de estudos avaliativos.

  3ieimpact.org

- Campbell Collaboration

  Biblioteca internacional de revisões sistemáticas sobre políticas sociais.

  campbellcollaboration.org

- What Works Network (Reino Unido)

  Plataformas governamentais com resumos de evidência sobre setores específicos de políticas públicas, como educação, saúde, economia, e segurança.

  gov.uk/guidance/what-works-network

- What Works Clearinghouse (EUA)

  Plataforma de evidências sobre o setor de educação:

  ies.ed.gov

Fonte: FGV CLEAR, elaboração própria.

Soluções baseadas em inteligência artificial estão transformando a forma como buscamos e organizamos evidências. Embora seu uso em avaliação de políticas públicas ainda seja relativamente recente, diversas aplicações já demonstram utilidade concreta — especialmente para revisões rápidas, exploração de literatura científica e síntese de achados. Com avanços em algoritmos de linguagem natural e integração com bancos de dados científicos, essas tecnologias ampliam o acesso a estudos relevantes e podem ser especialmente úteis para equipes que atuam com prazos curtos ou temas complexos. Exemplos de ferramentas e formas de uso estão apresentados no **Apêndice I**.

2.1.3 Atenção à qualidade da evidência

Nem todo estudo apresentado em um relatório, site institucional ou publicação técnica é, de fato, uma evidência confiável. Um estudo pouco robusto pode levar a conclusões equivocadas. E, quando essas conclusões embasam políticas públicas, os efeitos podem ser ineficácia, desperdício de recursos ou até danos à população.

Para reduzir esses riscos, é importante considerar alguns cuidados básicos ao selecionar evidências para apoiar a formulação de uma política:

- Verifique se o método utilizado é apropriado ao tipo de pergunta e intervenção analisada.
- Dê preferência a estudos revisados por pares ou publicados por instituições reconhecidas. Publicações acadêmicas, centros de pesquisa independentes

e organismos multilaterais costumam se preocupar com padrões de qualidade.

- Certifique-se de que há transparência metodológica, clareza nos procedimentos adotados e ausência de conflitos de interesse;
- Verifique se há corroboração por outras fontes — como revisões sistemáticas e meta-análises que reúnem e avaliam múltiplos estudos sobre o mesmo tema.

Um recurso útil é a abordagem da **pirâmide da evidência** (MURAD et al., 2016), amplamente usado na área da saúde, mas adaptável a outros campos. Nessa estrutura:

- No topo estão as revisões sistemáticas e meta-análises, que sintetizam estudos rigorosos com critérios transparentes;
- No meio, os estudos individuais com métodos experimentais ou quase-experimentais;
- Na base, evidências menos robustas, como estudos sem grupo de comparação, análises com baixa validade ou relatos baseados apenas em opinião.

O caso do uso do medicamento cloroquina durante a pandemia de COVID-19 ilustra bem a importância da qualidade das evidências. Estudos iniciais, frágeis e sem revisão por pares, sugeriram efeitos positivos que, posteriormente, foram refutados por pesquisas mais robustas. A mesma lógica se aplica às políticas públicas: quanto mais sólida a evidência utilizada, maiores as chances de desenhar uma intervenção eficaz, segura e bem direcionada.

Box 7 – Recursos úteis de organizações internacionais

Organizações internacionais, como o Banco Interamericano de Desenvolvimento (BID) e o Banco Mundial, oferecem uma variedade de recursos que sistematizam evidências sobre políticas públicas, representando fontes valiosas para quem formula, adapta ou aprimora programas governamentais.

O relatório do BID *Pequenos Empurrões para a América Latina e o Caribe* (2022) é um exemplo claro de como intervenções baseadas na ciência comportamental — os chamados *nudges* — vêm sendo aplicadas em diferentes países da região para resolver desafios como baixa arrecadação fiscal, evasão escolar, ineficiências em serviços públicos e desigualdades de gênero. O documento descreve as intervenções, e apresenta os resultados obtidos, com base em avaliações experimentais e quase-experimentais, mostrando o impacto concreto das mu-

danças de comportamento induzidas por essas políticas.

Já os *World Development Reports* (WDR), publicados anualmente pelo Banco Mundial, reúnem um vasto corpo de evidências internacionais sobre temas estratégicos. O WDR de 2018, por exemplo, revelou a existência de uma crise global de aprendizagem, defendendo o uso de avaliações educacionais e políticas baseadas em evidências para melhorar o aprendizado efetivo, e não apenas o acesso à escola. O WDR de 2019 abordou a transformação do mundo do trabalho em decorrência das novas tecnologias, com recomendações de políticas ativas de capacitação e proteção social para trabalhadores em transição.

Esses relatórios são úteis porque articulam evidências empíricas de múltiplos contextos com recomendações práticas, facilitando a adaptação local de soluções testadas e bem-sucedidas.

Fonte: FGV CLEAR, elaboração própria.

2.1.4 Aprender também com o que não funcionou

Avaliar alternativas exige olhar não apenas para o que funcionou, mas também para o que não funcionou — e por quê. Uma política pode falhar por má concepção, quando sua lógica causal é frágil ou incoerente com o diagnóstico. Mas também pode fracassar por falhas de implementação, mesmo que seu desenho seja promissor. Distinguir essas duas dimensões é fundamental: evita que políticas mal concebidas sejam replicadas e que boas ideias sejam descartadas injustamente por problemas de execução.

Ao analisar experiências anteriores, vale refletir sobre as seguintes questões:

- O problema de fundo era realmente semelhante ao enfrentado pela política em questão?
- A população-alvo tinha características comparáveis?
- A política foi bem implementada? Havia capacidade técnica e institucional?
- A avaliação foi rigorosa? Qual método foi utilizado?
- Quais fatores foram determinantes para o sucesso ou para o fracasso?

Essa análise crítica permite extrair aprendizados valiosos — não só para escolher melhor entre as alternativas, mas também para ajustar o desenho da política e antecipar riscos na sua aplicação.

2.1.5 E se a política for nova?

Nem toda política precisa — ou pode — replicar modelos existentes. Em muitos contextos, inovar é necessário: seja porque o problema enfrentado é inédito, seja porque as soluções já testadas não se aplicam ao contexto local. Nesses casos, a ausência de evidências prévias não deve paralisar a ação, mas exige **ainda mais cuidado no desenho da intervenção**.

Ao propor algo novo, é essencial garantir que a política seja **coerente com o diagnóstico** realizado, alinhada às causas identificadas e viável dentro das capacidades da equipe. É igualmente importante construir uma **lógica clara de transformação** — ou seja, uma Teoria da Mudança que explicite como as ações previstas devem levar aos resultados esperados.

Outro ponto central é incorporar desde o início um **plano de monitoramento e avaliação**, que permita acompanhar a implementação, aprender com os resultados iniciais e ajustar o percurso, se necessário. Sempre que possível, recomenda-se **testar a política em pequena escala (piloto)** antes da ampliação. Isso permite avaliar a viabilidade prática, mensurar efeitos preliminares e fazer ajustes antes de investir em larga escala.

Inovação em políticas públicas é desejável — mas deve ser feita com responsabilidade, planejamento e disposição para aprender ao longo do caminho

2.1.6 Quando há várias alternativas possíveis: como escolher?

A análise de evidências frequentemente revela que há mais de uma alternativa viável para enfrentar um mesmo problema público. Cada opção pode atuar sobre diferentes causas, envolver níveis variados de complexidade, exigir graus distintos de articulação institucional e produzir efeitos em ritmos e dimensões diversas. Em geral, todas têm algum potencial de impacto — mas nem sempre é possível implementar todas ao mesmo tempo.

É nesse ponto que a formulação exige uma escolha estratégica, fundamentada em critérios objetivos e sustentada por evidências. A decisão sobre qual política (ou combinação de políticas) seguir não deve se basear apenas em intuições ou conveniências, mas em uma análise comparativa estruturada, que considere múltiplas dimensões do desenho e da viabilidade da intervenção.

Um conjunto clássico de critérios para orientar essa escolha é proposto por Kraft e Furlong (2019), alinhado às melhores práticas internacionais e nacionais recomendadas pelo Green Book (Tesouro Nacional Britânico, Reino Unido) e pelo Tribunal de Contas da União (TCU). No **Box 8** a seguir, resumimos cada um desses critérios com orientações práticas para sua aplicação.

Box 8 – Como comparar alternativas de políticas públicas

Critério	Definição	Como aplicar na prática
Efetividade	Capacidade da política de atingir seus objetivos declarados.	Pergunte: "Essa alternativa resolve o problema?" Avalie resultados esperados com base em evidências.
Eficiência	Relação entre os custos envolvidos e os benefícios gerados.	Compare o custo por beneficiário e por resultado, ou use análises de custo-benefício ou custo-efetividade. Mapeie claramente todos os custos para fortalecer o planejamento orçamentário e a transparência. Avalie também o impacto fiscal, custos de implementação, manutenção e monitoramento.
Equidade	Como os custos e benefícios são distribuídos entre diferentes grupos.	Considere se a política reduz desigualdades ou tende a beneficiar desproporcionalmente grupos privilegiados.
Liberdade	Grau de liberdade ou restrição que a política impõe às pessoas.	Avalie se a política restringe direitos, impõe obrigações ou interfere em escolhas individuais.
Viabilidade política	Nível de apoio ou resistência entre atores políticos relevantes. Avalie os arranjos institucionais e o ambiente político no qual a política será inserida.	Identifique atores influentes, suas motivações favoráveis ou contrárias, disponibilidade de recursos (financeiros, simbólicos ou institucionais) e riscos de bloqueio.
Aceitação social	Grau de apoio ou oposição da população à política proposta.	Verifique percepções sociais, alinhamento com valores culturais e resultados de consultas públicas. Avalie riscos de rejeição ou baixa adesão.
Viabilidade técnica	Existência de conhecimentos, capacidades e tecnologias para executar a política.	Analise se há capacidade técnica instalada ou necessidade de formação, apoio externo ou inovação.
Viabilidade administrativa	Capacidade da estrutura administrativa de implementar e coordenar a política.	Avalie se os órgãos envolvidos têm recursos, processos e articulação institucional adequados.

Fonte: FGV CLEAR, elaboração própria.

Esses critérios ajudam a estruturar o debate sobre alternativas, promovendo transparência, coerência e robustez técnica no processo de formulação. Além disso, permitem dialogar com diferentes públicos — técnicos, gestores, políticos, sociedade civil — a partir de um vocabulário comum.

Importante ressaltar que não existe um critério absoluto que prevaleça sobre os demais. Em muitos casos, a política mais eficaz pode ser também a mais difícil de implementar; uma opção com forte apoio social pode ser menos eficiente que

outra com menor visibilidade pública. Por isso, a decisão final exige equilíbrio e julgamento técnico-político.

A abordagem recomendada pelo *Green Book*, do Reino Unido, e pelo Tribunal de Contas da União reforça esses pontos, destacando ainda a importância da análise aprofundada dos riscos e incertezas associados às alternativas. Essa análise envolve uma identificação clara de riscos potenciais, incluindo fatores econômicos, sociais, políticos, técnicos e ambientais. É necessário avaliar a probabilidade de ocorrência desses riscos e seu possível impacto sobre a política proposta, além de definir estratégias específicas para sua mitigação e gestão.

Exemplo prático 9— Escolhendo entre alternativas no caso da baixa frequência escolar

Retomando o exemplo da baixa frequência escolar, o diagnóstico revelou uma série de causas diretas relevantes: ausência de alimentação escolar, transporte escolar insuficiente, existência de trabalho infantil, condições precárias de infraestrutura escolar, problemas de saúde não tratados e absenteísmo docente. Todas essas causas estavam conectadas a fatores estruturantes, como pobreza persistente das famílias, desigualdade territorial no acesso a serviços, baixo investimento público em educação básica e ausência de políticas de proteção à infância.

Com base nesse diagnóstico, diversas alternativas de política foram consideradas para enfrentar o problema:

- Transferência de renda condicionada à frequência escolar, para aliviar a necessidade de trabalho infantil e incentivar a assiduidade;
- Oferta de alimentação escolar reforçada, para reduzir custos indiretos das famílias e tornar a escola mais atrativa;
- Transporte gratuito e busca ativa, com foco em regiões de maior vulnerabilidade;
- Ações intersetoriais com saúde e assistência social, garantindo atendimento a crianças com problemas de saúde ou expostas à violência;
- Intervenções na escola, como melhoria da infraestrutura, reforço escolar e valorização dos professores.

Cada alternativa apresentava lógicas de atuação distintas, incidindo sobre causas específicas e exigindo diferentes níveis de articulação e investimento. Para orientar a escolha, a equipe técnica aplicou uma análise estruturada com base em critérios clássicos de formulação, como os propostos por Kraft e Furlong (2019), avaliando:

- Efetividade: qual alternativa tinha maior potencial de aumentar a frequência escolar?
- Eficiência: qual gerava maior impacto por real investido?
- Equidade: quais estratégias beneficiariam de forma mais direta os alunos em situação de maior vulnerabilidade?
- Viabilidade técnica e administrativa: havia capacidade instalada para executar as ações propostas?
- Viabilidade política e aceitação social: quais alternativas teriam mais apoio entre gestores, escolas e comunidades?

Ao final dessa análise comparativa, duas estratégias se destacaram como mais promissoras:

- A transferência de renda condicionada à frequência escolar, reconhecida na literatura internacional (SNILSTVEIT ET AL., 2015) como uma das intervenções com maior impacto comprovado sobre matrícula e assiduidade em países de baixa e média renda;
- A oferta de alimentação escolar reforçada, também respaldada por evidências robustas (SNILSTVEIT ET AL., 2015), especialmente em contextos marcados por pobreza, onde a dificuldade de acesso à alimentação adequada compromete a frequência regular.

No entanto, ao considerar a viabilidade institucional e legal, os formuladores verificaram que o Brasil já conta com o Programa Bolsa Família, que inclui como condicionalidade a frequência mínima escolar. Como o estado em questão não tinha margem orçamentária, nem atribuição legal, para criar uma política de transferência de renda, decidiu-se priorizar uma alternativa complementar, viável e de rápida implementação.

A escolha recaiu sobre a expansão e qualificação da alimentação escolar, com foco em regiões mais vulneráveis. Embora o Brasil conte com o Programa Nacional de Alimentação Escolar (PNAE), a equipe identificou limitações na execução local — como falta de regularidade na oferta, cardápios inadequados e cobertura insuficiente nas escolas de maior vulnerabilidade social. Assim, lançaram o Programa de Reforço da Alimentação Escolar.

A medida foi considerada efetiva, equitativa, tecnicamente viável e com forte apoio social, especialmente entre os estudantes mais afetados pelas causas diretas diagnosticadas — como a ausência de alimentação escolar, o trabalho infantil e a pobreza persistente em geral. A

decisão foi tomada com base em evidências robustas e no uso de critérios técnicos para comparação entre alternativas, favorecendo uma escolha informada, realista e ancorada na capacidade de ação do governo local.

Fonte: FGV CLEAR, elaboração própria com base em Kraft e Furlong (2019) e Snilstveit et al. (2015).

Com base no diagnóstico e na comparação cuidadosa entre alternativas a partir dos critérios apresentados, a equipe gestora consegue definir qual política — ou combinação de políticas — será adotada para enfrentar o problema público. A partir dessa escolha, o próximo passo é **tornar explícita a lógica que conecta as ações propostas aos resultados esperados**. Essa representação estruturada é feita por meio da **Teoria do Programa e Teoria da Mudança**, que será o foco da próxima seção.

Aplicando na prática

- Busque exemplos em fontes confiáveis: repositórios de evidência, avaliações, estudos técnicos e especialistas da área.
- Verifique se o problema enfrentado em outros contextos é realmente comparável ao seu — o contexto importa.
- Priorize alternativas com evidência robusta de impacto e aderência ao seu território.
- Considere também o que não funcionou: isso ajuda a evitar armadilhas de desenho ou execução.
- Se não houver evidência suficiente, considere simular cenários ou testar em pequena escala (pilotos) antes da implementação em larga escala.
- Compare as opções com base em critérios estruturados, como efetividade, eficiência, equidade, viabilidade técnica, política e administrativa
- Lembre-se: a melhor alternativa nem sempre é a mais ambiciosa, mas sim a mais aderente ao contexto, às capacidades locais e ao momento político.
- Registre o raciocínio por trás da escolha: isso será útil para o desenho da Teoria da Mudança e para futuras avaliações.

3 Estruturando a lógica da política: a Teoria do Programa e a Teoria da Mudança[7]

A formulação de políticas públicas exige mais do que boas intenções ou ações pontuais: é necessário estruturar uma **lógica de intervenção clara**, com base em evidências, que explique por que uma determinada combinação de ações, recursos e entregas deve gerar os resultados pretendidos. Essa lógica é chamada de **Teoria do Programa** — e sua representação prática mais difundida é a T**eoria da Mudança (TdM)**

A Teoria da Mudança descreve, de forma visual, o caminho que conecta o problema identificado aos impactos desejados, passando pelos principais elementos de implementação: insumos, atividades, produtos e resultados. Ela também explicita as **hipóteses causai**s que sustentam esse caminho e os riscos que podem ameaçar sua implementação (FUNNELL & ROGERS, 2011).

A **Teoria do Programa** pode ser expressa de forma sintética por meio de uma narrativa que explicita o raciocínio causal da intervenção. Um formato útil para isso é:

Se fizermos [atividades com base em insumos], então entregaremos [produtos], o que levará a [resultados esperados], contribuindo para [impactos desejados].

Exemplo prático 10— Teoria do Programa do Programa de Reforço da Alimentação Escolar- PRAE

Se reforçarmos o planejamento nutricional, a aquisição de alimentos saudáveis e a preparação diária das refeições nas escolas, com base em infraestrutura adequada e equipe capacitada, então entregaremos refeições balanceadas e regulares, alinhadas às necessidades dos alunos, o que levará a uma maior disposição e motivação para frequentar a escola e maior frequência escolar, contribuindo para a redução da evasão e melhoria dos indicadores educacionais e nutricionais.

Fonte: FGV CLEAR, elaboração própria.

[7] A literatura na qual este capítulo se baseia é constituída pelos trabalhos de Brasil (2018a), Comissão Europeia (2024), Funnell e Rogers (2011), IJSN (2018a), OCDE (2002; 2010), Roberts e Khattri (2012), Rogers (2008), Weiss (1997).

3.1 O que é a Teoria da Mudança?

A **Teoria da Mudança (TdM)** é uma tradução visual da Teoria do Programa. Ela explicita como a política proposta espera transformar recursos em efeitos positivos para a sociedade, a partir de um encadeamento causal estruturado.

Esse encadeamento parte de uma lógica simples, mas poderosa: se realizarmos determinadas atividades, entregaremos produtos; esses produtos gerarão resultados, que contribuirão para alcançar os impactos desejados. A TdM torna explícita essa lógica, conectando cada elo da intervenção com clareza e coerência.

Mais do que uma representação gráfica, a TdM orienta decisões, alinha expectativas e cria as bases para o monitoramento e a avaliação. Sempre que possível, deve ser **fundamentada em evidências**, seja de experiências anteriores ou literatura especializada.

A TdM é o mapa da política: mostra aonde se quer chegar, por qual caminho, com quais recursos e o que se espera alcançar em cada etapa. Por este motivo, a TdM é uma importante aliada em todos os momentos do ciclo da política pública.

Box 9 – Uso da TdM ao longo do ciclo da política pública

Momento do ciclo	Uso da Teoria da Mudança
Ex ante	Planejar com base em lógica causal robusta, explicitar premissas e selecionar indicadores e metas para planejar de forma adequada o plano de monitoramento e avaliação.
Implementação	Guiar a gestão por resultados, monitorar a execução e identificar desvios.
Ex post	Verificar se a política funcionou como previsto, apoiar a elaboração de perguntas avaliativas para distinguir falhas de teoria e de implementação.

Fonte: FGV CLEAR, elaboração própria.

3.2 Os componentes da Teoria da Mudança

A **Teoria da Mudança (TdM)** estrutura a política pública em uma cadeia lógica organizada em cinco dimensões, cada uma desempenhando funções específicas e essenciais para o sucesso da intervenção. Essa estrutura permite visualizar claramente a sequência lógica entre insumos, atividades, produtos, resultados e impactos esperados.

Além disso, a construção da TdM inclui explicitamente a identificação de **hipóteses** e **riscos** relacionados à implementação da política pública, essenciais para entender as condições necessárias para o alcance dos objetivos almejados.

3.2.1 Insumos: O que é necessário mobilizar para que a política exista?

Os insumos são os recursos que viabilizam a implementação da política. Incluem:

- Orçamento (federal, estadual, municipal, internacional)
- Recursos humanos (equipe técnica, servidores, parceiros)
- Infraestrutura física (espaços, equipamentos, tecnologia)
- Bases legais e normativas
- Sistemas e ferramentas de gestão

⇒ *A ausência ou inadequação dos insumos compromete diretamente a realização das atividades previstas.*

3.2.2 Atividades: O que será feito com os insumos mobilizados?

As atividades são as ações concretas que a equipe executa para colocar a política em prática. Dividem-se entre:

- **Atividades-meio**: necessárias para viabilizar o funcionamento do programa (ex: capacitar equipe, cadastrar beneficiários, contratar fornecedores)
- **Atividades-fim**: diretamente voltadas ao público-alvo (ex: ministrar oficinas, realizar visitas domiciliares, oferecer serviços)

⇒ *As atividades precisam ser factíveis, bem descritas, expressas a partir de verbos no infinitivo e claramente conectadas aos produtos esperados.*

3.2.3 Produtos: O que será entregue diretamente ao público-alvo?

Os produtos são os bens ou serviços produzidos pelas atividades da política. São efeitos imediatos e tangíveis. Devem ser mensuráveis e estar sob controle direto da equipe de implementação.

Exemplos:

- Famílias contempladas com o benefício de transferência de renda.
- Professores com o certificado de conclusão do curso de capacitação.
- Famílias atendidas regularmente através das visitas domiciliares.
- Refeições servidas

⇒ *Produtos são entregas concretas e expressas no particípio passado. Apesar de relevantes, não necessariamente garantem os resultados esperados.*

3.2.4 Resultados: O que deve mudar no público-alvo em decorrência do uso dos produtos?

Os resultados representam mudanças de comportamento, conhecimento, atitude ou condição das pessoas beneficiadas. Podem ser:

- **De curto prazo**: Por exemplo, aumento na frequência escolar, melhor conhecimento sobre alimentação saudável
- **De médio prazo**: Por exemplo, melhor desempenho acadêmico, adoção de práticas parentais positivas

⇒ *Os resultados são influenciados pela política, mas não são completamente controláveis — eles dependem de outras variáveis e contextos.*

3.2.5 Impactos: Quais são as transformações estruturais esperadas no longo prazo?

O impacto é o objetivo final da política — a transformação de longo prazo que se pretende atingir. Está diretamente relacionado à solução (ou mitigação) do problema público identificado.

Exemplos:

- Redução da pobreza
- Aumento da escolaridade média
- Melhoria da saúde materno-infantil
- Redução da violência

⇒ *O impacto é a razão de existir da política. Mas ele depende de muitas outras forças além da intervenção isolada.*

Box 10 – Os elementos da Teoria da Mudança

A TdM é composta por cinco elementos principais, organizados em cadeia causal. Cada nível depende do anterior e pressupõe condições para sua efetivação:

Componente	Definição	Exemplos práticos
Insumos	Recursos necessários para viabilizar a política. Incluem recursos financeiros, humanos, técnicos, institucionais e legais.	Equipe gestora, orçamento, marcos normativos, infraestrutura, sistemas de informação.
Atividades	Ações realizadas com os insumos. Transformam recursos em bens ou serviços. Podem ser diretas (voltadas ao público) ou indiretas (de apoio).	Cadastrar beneficiários, processar pagamentos, treinar profissionais, mobilizar parceiros.
Produtos	Bens e serviços entregues à população como resultado imediato das atividades. São mensuráveis e sob controle direto da política.	Famílias atendidas, refeições servidas, professores formados, bolsas pagas.
Resultados	Mudanças observadas nos beneficiários diretos: em comportamento, condição ou percepção. São efeitos de curto ou médio prazo.	Crianças com maior frequência escolar, jovens com mais habilidades, aumento da renda disponível.
Impacto	Transformações de longo prazo e de maior abrangência na sociedade, relacionado ao problema que motivou a criação da política. A política contribui para eles, mas não os determina isoladamente.	Redução da evasão escolar, aumento da escolaridade média, redução da pobreza.

Fonte: Elaboração própria com base em Funnell e Rogers (2011); Roberts e Khattri (2012).

3.2.6 Hipóteses e riscos: os elos invisíveis da cadeia

Uma Teoria da Mudança bem construída não se limita a listar atividades e resultados esperados — ela também explicita **os fatores que sustentam a lógica da intervenção e os elementos que podem comprometer sua execução**. Entre esses fatores estão os riscos e as **hipóteses**, que funcionam como **elos invisíveis** da cadeia de resultados. Identificá-los é essencial para fortalecer o planejamento, o monitoramento e a avaliação da política pública.

Neste guia, adotamos uma distinção conceitual que tem se mostrado especialmente útil para gestores e avaliadores:

- Riscos (de implementação): ameaças que podem comprometer a realização de etapas previstas da política, especialmente nos elos iniciais da cadeia de

resultados (de insumos a atividades, e de atividades a produtos). São situações que podem impedir a entrega dos bens e serviços planejados, como atrasos operacionais, falhas logísticas ou problemas de gestão.

- Hipóteses (de efetividade): pressupostos que precisam se confirmar para que os produtos entregues pela política gerem os efeitos esperados — resultados e impactos. Dizem respeito à lógica causal da intervenção e, em geral, envolvem fatores externos ao controle direto da gestão.

Tornar explícitos riscos e hipóteses permite à gestão pública antecipar fragilidades e ajustar estratégias de implementação. Também contribui para estruturar perguntas orientadoras que apoiarão a avaliação da política ao longo do ciclo.

⇒ **IMPORTANTE:** Identificar os riscos ajuda a prevenir falhas na entrega. Tornar visíveis as hipóteses permite testar se a política está funcionando como se espera. A atenção a ambos deve orientar o desenho de intervenções mais robustas e com maior potencial de impacto.

Box 11 – O que diz a literatura SOBRE RISCOS E HIPÓTESES?

A distinção entre riscos e hipóteses não é unívoca na literatura. Em algumas abordagens de organismos multilaterais, como a OCDE e o Banco Mundial, os riscos e hipóteses são considerados como fatores externos: riscos são eventos adversos que podem comprometer a política; já as hipóteses (ou pressupostos) são condições que se espera que estejam presentes para que a intervenção funcione (ROBERTS & KHATTRI, 2012; OCDE, 2002; 2010). A diferença, nesses casos, está sobretudo na forma de enunciação — negativa no caso dos riscos, positiva no caso das hipóteses.

Autores relevantes na literatura de Teoria do Programa, como Weiss (1997) e Rogers (2008), definem as hipóteses como explicações causais que articulam os elos da intervenção — ou seja, por que se espera que um produto leve a um resultado, ou um resultado a um impacto.

Apesar das variações conceituais, há um ponto de convergência entre as diferentes tradições: explicitar as suposições que sustentam a lógica da política e identificar os riscos que podem comprometer sua execução são práticas essenciais para o planejamento, a gestão e a avaliação. A escolha metodológica adotada neste guia — distinguir entre riscos de implementação e hipóteses de efetividade — busca justamente apoiar uma prática mais clara, analítica e orientada para a tomada de decisão pública.

Fonte: FGV CLEAR, elaboração própria com base em Roberts & Khattri, (2012), OCDE (2002; 2010), Weiss (1997) e Rogers (2008).

3.3 Como construir a Teoria da Mudança na prática?

Existem diferentes ferramentas que buscam representar a Teoria do Programa de uma política pública, como o Marco Lógico, o Quadro Lógico, o Modelo Lógico, a Cadeia de Resultados, a própria Teoria da Mudança (TdM), entre outras. Todas têm o objetivo de explicitar como uma intervenção pretende alcançar os efeitos esperados.

Contudo, neste guia, optamos por utilizar a TdM por ser a abordagem mais adequada para construir, de forma crítica e participativa, o raciocínio causal completo da política.

Diferentemente de ferramentas mais lineares e descritivas, a TdM:

- Permite representar múltiplos caminhos causais e não apenas cadeias lineares;
- É mais flexível e adaptável a cada contexto, podendo ser construída de forma visual, narrativa ou combinada;
- Valoriza o pensamento crítico, ao explicitar as hipóteses que sustentam cada ligação causal e os riscos que podem comprometer o funcionamento da política;
- Favorece a análise ex ante e ex post, sendo útil para planejar, monitorar e revisar a intervenção ao longo do tempo.

Sua representação se dá por meio de blocos encadeados por setas, que simbolizam as relações de causa e efeito entre os componentes da intervenção. Essas setas não são meros conectores visuais: elas carregam suposições que precisam ser plausíveis, baseadas em evidência ou lógica teórica.

⇒ *Uma pergunta-chave ao construir a TdM é: "Qual atividade (ou conjunto de atividades) gera determinado produto? E qual produto (ou combinação de produtos) é necessário para produzir determinado resultado?"*

É fundamental compreender as relações entre **atividades**, **produtos** e **resultados** ao construir uma TdM. De modo geral, toda atividade da política deve gerar pelo menos um produto – uma entrega concreta, mensurável e observável. Afinal, ações sem produtos verificáveis indicam fragilidade no desenho da intervenção.

Contudo, nem sempre a relação entre atividade e produto é direta. Em muitos casos, um mesmo produto resulta da **combinação de diversas atividades** articuladas. É o que ocorre, por exemplo, na implementação de um serviço público que exige múltiplas ações coordenadas, como a contratação de equipe, a aquisição de insumos e a realização de capacitações. Embora distintas, essas atividades convergem para uma única entrega estruturada e funcional.

O mesmo raciocínio vale para os resultados. Em políticas públicas mais complexas, os efeitos esperados raramente decorrem de uma entrega isolada. Frequentemente, é a interação **entre diferentes produtos** – como a infraestrutura implantada, os serviços ofertados e os mecanismos de acompanhamento – que permite alcançar os resultados pretendidos.

Portanto, ao mapear a lógica da intervenção, é preciso reconhecer que as ligações entre atividades, produtos e resultados podem assumir formatos variados: de relações simples "um para um" a configurações mais integradas "muitos para um", sempre de acordo com os objetivos e a complexidade da política analisada.

No **Box 12** a seguir, propomos um passo a passo para a construção da TdM, com base nas práticas do FGV CLEAR.

Box 12 – Etapas para construir uma Teoria da Mudança robusta:

1. Reafirme o problema público e o impacto desejado
    - Qual é o problema a ser enfrentado?
    - Qual transformação estrutural representaria o sucesso da política?
    - O impacto desejado responde ao problema?

2. Mapeie os resultados necessários para alcançar o impacto
    - Quais mudanças no comportamento, atitude ou condição do público-alvo são essenciais para gerar o impacto?
    - Esses resultados podem ser organizados em curto, médio e longo prazo?

3. Liste os produtos a serem entregues
    - Quais bens ou serviços serão entregues diretamente ao público?
    - Quais produtos são necessários (individualmente ou em conjunto) para gerar os resultados esperados?

4. Descreva as atividades necessárias
    - Quais ações devem ser realizadas para produzir cada produto?
    - Há atividades que contribuem simultaneamente para mais de um produto?

- Há atividades de apoio (ex: gestão, articulação intersetorial)?

Quando apropriado, agrupe atividades que, em conjunto, são necessárias para a entrega de um produto relevante.

5. Relacione os insumos requeridos

- Quais recursos financeiros, humanos, materiais, tecnológicos ou normativos são necessários para viabilizar as atividades?
- Há capacidade institucional para mobilizar esses insumos?

6. Construa o diagrama da TdM

- Organize os cinco elementos principais da cadeia:
- Insumos → Atividades → Produtos → Resultados → Impacto
- Use setas para representar a lógica de causalidade entre os elementos.
- Indique claramente:
  - Quais atividades criam quais produtos;
  - Quais produtos contribuem para quais resultados;
  - Quais resultados são necessários para alcançar o impacto.

7. Formule as hipóteses e identifique os riscos

- O que estamos pressupondo em cada seta causal? Essas suposições são plausíveis?
- Há evidência empírica ou base teórica que sustente essas hipóteses?
- Quais fatores podem comprometer a realização das atividades, a entrega dos produtos ou a ocorrência dos resultados?
- Como mitigar os riscos mais relevantes?

8. Valide com diferentes públicos

- Submeta a TdM a quem conhece o programa na prática: gestores, técnicos, parceiros, beneficiários.

- O encadeamento faz sentido?
- Há elementos esquecidos ou hipóteses frágeis?

9. Ajuste, refine, documente

- Elabore uma versão final visual da TdM (diagrama) e uma narrativa explicativa que detalhe o raciocínio por trás de cada etapa.
- Guarde versões iterativas para revisões futuras (especialmente útil em avaliações durante a execução da política que podem apoiar redesenhos).

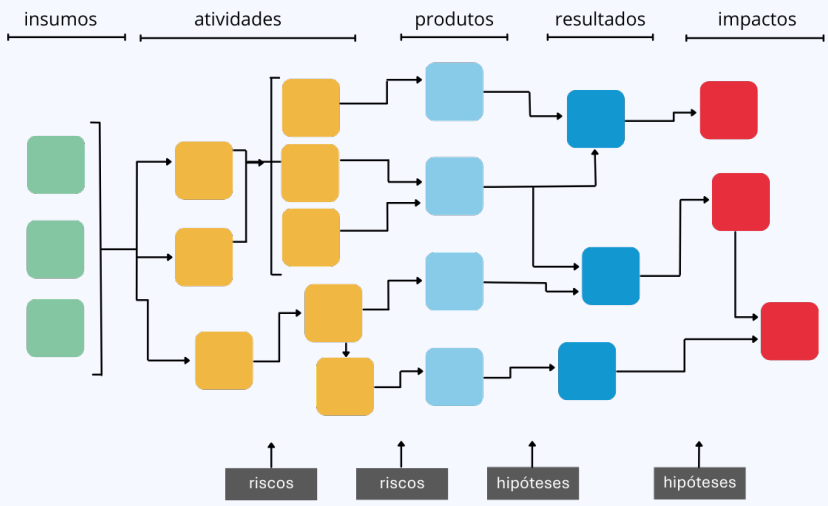
Nota: A ordem sugerida é ideal quando se utiliza a TdM para desenhar ou redesenhar uma política pública, partindo claramente do impacto desejado. No entanto, em situações em que a política já está em andamento ou as ações estão pré-definidas, pode ser mais apropriado iniciar pelos insumos e atividades existentes, ajustando progressivamente a cadeia até que se alcance o impacto pretendido.

Fonte: FGV CLEAR, elaboração própria.

3.4 Representação visual: o diagrama da Teoria da Mudança

A representação gráfica da TdM ajuda a comunicar com clareza a lógica do programa e suas principais relações causais. O diagrama abaixo ilustra, de forma simplificada, o encadeamento de **insumos → atividades → produtos → resultados → impacto**, com ênfase nas setas que mostram o que leva a quê, e com a sinalização das **hipóteses críticas e riscos mapeados**. A **Figura 8** explicita um dos modelos que podem ser utilizados para representar essa estrutura.

Figura 8: Representação gráfica de uma Teoria da Mudança



Fonte: FGV CLEAR, elaboração própria.

Esse tipo de representação pode ser complementado com uma narrativa explicativa da cadeia causal, que detalhe a lógica de cada seta e as condições para que cada transição ocorra.

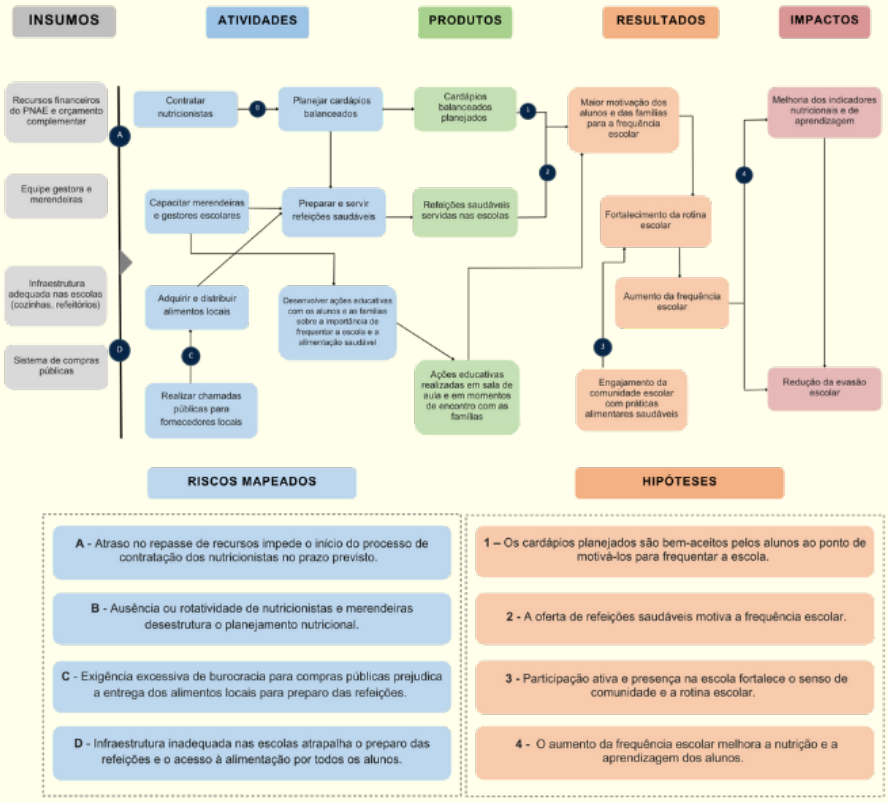
Exemplo prático 11— Teoria da Mudança para o Programa de Reforço da Alimentação Escolar

Problema público:

Baixa frequência escolar e alta evasão no ensino fundamental em áreas vulneráveis.

Impacto desejado:

Aumento da frequência e redução da evasão escolar e melhoria dos indicadores nutricionais e de aprendizagem dos alunos.



Observação didática:

A relação entre atividades e produtos pode seguir diferentes arranjos:

- Uma atividade específica (como planejar cardápios balanceados) pode gerar um produto direto (cardápios balanceados planejados);
- Um produto mais complexo (ex: refeições saudáveis servidas nas escolas) pode depender de várias atividades combinadas (compra, capacitação, preparo, serviço) ;
- É possível também que um resultado esperado dependa da interação entre múltiplos produtos, não havendo necessariamente a correspondência de 1:1 (ex: maior motivação dos alunos e das famílias para a frequência escolar).

Fonte: FGV CLEAR, elaboração própria.

3.5 Identificando falhas: teoria ou implementação?

A Teoria da Mudança também é uma ferramenta útil para analisar resultados insatisfatórios e diagnosticar porque uma política pública não produziu os efeitos

esperados. Ao explicitar a lógica causal que conecta insumos, atividades, produtos, resultados e impactos, a ferramenta permite identificar dois tipos centrais de falha:

3.5.1 Falha de implementação

A falha na implementação acontece quando a política foi **mal executada**, e por isso os resultados não se materializaram.

- As atividades não foram realizadas conforme o planejado;
- Os produtos não foram entregues ou foram entregues com baixa qualidade;
- Faltaram insumos, houve atrasos ou problemas operacionais.

Exemplo: os cardápios foram planejados por nutricionistas, mas os alimentos não chegaram às escolas por falhas na logística. As refeições não foram servidas e a política fracassou por falha de execução.

Figura 9: Falha na implementação



Fonte: FGV CLEAR, elaboração própria.

3.5.2 Falha na teoria (ou falha no desenho)

A falha na teoria acontece quando a política foi bem executada, mas a **lógica da intervenção não se sustentou** — ou seja, os produtos não foram capazes de gerar os resultados esperados.

- As hipóteses causais eram frágeis ou incorretas;
- Os produtos entregues não eram suficientes ou não surtiam o efeito esperado sobre o comportamento dos beneficiários;
- Os resultados dependiam de fatores externos não considerados no desenho.

Exemplo: as refeições foram servidas conforme o planejado, mas a frequência escolar não aumentou porque o principal obstáculo era o transporte escolar, não a alimentação.

Distinguir essas duas falhas é **fundamental para a correção da política**. Corrigir uma falha de implementação exige melhorias operacionais; já uma falha de teoria demanda uma revisão da lógica do programa, suas hipóteses e caminhos causais, podendo levar ao redesenho da política.

⇒ *Uma política pode ser bem implementada e ainda assim fracassar — se estiver mal desenhada. E pode ser bem desenhada, mas falhar na prática — se for mal implementada.*

Figura 10: Falha na teoria (ou no desenho)



Fonte: FGV CLEAR, elaboração própria.

4 Estratégias de Mitigação de Riscos: antecipar problemas para proteger os resultados[8]

Toda política pública está sujeita a incertezas. Mesmo com um desenho bem estruturado, condições imprevistas, falhas operacionais, resistências institucionais ou mudanças externas podem comprometer os resultados pretendidos. Por isso, **antecipar riscos e planejar estratégias de mitigação é parte fundamental do processo de formulação**.

Ao construir a Teoria da Mudança (TdM), já mapeamos os **principais riscos** associados a cada elo da cadeia causal. Essa identificação não deve ficar apenas no plano do diagnóstico. Cada risco relevante precisa ser acompanhado de uma **estratégia clara de mitigação**, ou seja, de ações preventivas, corretivas ou adaptativas capazes de reduzir a probabilidade de ocorrência do risco ou minimizar seus efeitos caso ele se concretize.

4.1 Como os riscos se conectam à avaliação?

Na prática, **cada risco identificado na TdM é também uma suposição de fragilidade** que deve ser monitorada e pode originar **uma pergunta avaliativa de implementação**. Isso permite não apenas saber o que deu certo ou errado, mas também por que e sob quais condições.

Exemplo prático 12— Programa de Reforço da Alimentação Escolar

- Risco identificado: falta de fornecedores locais aptos ou logística ineficiente compromete a entrega dos alimentos.
- Mitigação proposta: contratação descentralizada por escola + reserva técnica de alimentos.
- Pergunta avaliativa derivada: A entrega da merenda foi regular? Os atrasos afetaram a frequência dos alunos?

Fonte: FGV CLEAR, elaboração própria.

[8] Este capítulo está baseada na literatura técnica produzida por BetterEvaluation (s.d.), Comissão Europeia (2024), Imas e Ray (2009), Imas e Rist (2009), Kusek e Rist, (2004), Reino Unido (2022), UNFPA (2024).

4.2 Etapas para elaborar estratégias de mitigação

Recomendamos as seguintes etapas para a elaboração de uma estratégia de mitigação:

1. Retomar a Teoria da Mudança e revisar os riscos identificados para cada elo;

Para cada risco relevante, classificar sua:

- Probabilidade (alta, média, baixa);
- Impacto potencial sobre os resultados (alto, médio, baixo);

2. Selecionar os riscos com maior criticidade (alta probabilidade ou alto impacto);

Definir, para cada um deles:

- Ações preventivas: o que pode ser feito para evitar o risco?
- Ações de contingência: o que fazer se o risco se materializar?
- Responsáveis e recursos necessários para aplicar essas medidas.

Exemplo prático 13— Programa de Reforço da Alimentação Escolar: estratégias de mitigação de riscos

Risco identificado	Elo da TdM	Probabilidade de ocorrência	Impacto	Estratégia de mitigação	Responsável
Atraso no repasse de recursos	Insumo	Alta	Alto	Garantir dotação plurianual; acordos formais com o tesouro; cláusula de liberação automática parcial	Secretaria de Finanças
Falta de merendeiras nas escolas	Atividade	Média	Médio	Cadastro, reserva de profissionais; terceirização parcial	Coordenação local
Superlotação de turmas por aumento da demanda	Produto	Baixa	Alto	Mecanismo de triagem inicial com fila de espera rotativa; expansão escalonada por território	Secretaria de Educação

Fonte: FGV CLEAR, elaboração própria.

A estratégia de mitigação de riscos deve ser um **instrumento vivo**, vinculado ao monitoramento da política. Para cada risco crítico, o plano deve prever:

- **Indicadores de alerta precoce** (ex: número de escolas sem merenda por semana);
- **Fluxo de resposta rápida** (quem deve ser acionado, em quanto tempo, com quais recursos);

- **Documentação e registro dos eventos** para aprendizado institucional;
- **Acompanhamento periódico em reuniões de gestão intersetoriais**.

Políticas públicas bem formuladas não são imunes ao risco — mas são capazes de **antecipar vulnerabilidades, se adaptar com agilidade e proteger seus objetivos centrais**. A mitigação de riscos é parte fundamental da coerência da política e um indicador de maturidade da gestão pública.

5 Avaliação de Desenho: Análise Crítica da Teoria da Mudança[9]

5.1 O que é a avaliação de desenho?

A avaliação de desenho é uma etapa fundamental da gestão baseada em evidências. Seu objetivo é examinar criticamente a lógica interna de um programa ou intervenção, idealmente antes da sua implementação. Diferentemente das avaliações de resultados e de impacto, que focam nos efeitos alcançados, a avaliação de desenho concentra-se na teoria por trás do programa — ou seja, verifica se a sequência de etapas planejadas faz sentido lógico, está embasada em evidências e é compatível com o contexto de implementação da política. Em outras palavras, trata-se de analisar criticamente a Teoria da Mudança (TdM), identificando eventuais fragilidades, bem como riscos e hipóteses questionáveis. Essa abordagem ajuda gestores e avaliadores a ajustarem o caminho antes que recursos sejam investidos em estratégias possivelmente mal concebidas.

Avaliar o desenho da política é, portanto, um exercício que permite antecipar falhas, ajustar trajetórias e garantir que a política tenha uma base lógica sólida antes de ser colocada em prática. É também uma ferramenta de aprendizado institucional, pois ajuda equipes a refletirem criticamente sobre suas suposições, recursos e estratégias. A pergunta-chave da avaliação de desenho é: Essa política tem condições de funcionar?

5.2 Quando aplicar a avaliação de desenho?

A avaliação de desenho pode ser útil em dois momentos distintos:

- ***Ex ante***: antes da implementação. Esse é o momento ideal, ajudando a aprimorar o desenho e prevenir falhas de concepção.
- ***Ex post***: após a implementação. Recomendada principalmente quando os resultados esperados não foram alcançados. Permite analisar se o fracasso decorre de falhas na teoria do programa, e não apenas na execução.

Em ambos os casos, a avaliação funciona como uma ferramenta que testa a plausibilidade da intervenção e orienta ajustes estratégicos no desenho.

[9] Este capítulo se baseia na literatura produzida por Brasil (2018a), Funnell e Rogers (2011), IJSN (2018a), Pawson e Tilley (1997), OCDE (2002; 2010; 2023), Roberts e Khattri (2012), Rogers (2008), Rossi et al. (2004) , Sánchez (2024), Vaessen (2017), Weiss (1997).

5.3 Dimensões centrais da avaliação de desenho

A avaliação de desenho examina três dimensões principais:

- **Coerência interna**: a lógica causal da TdM faz sentido? Cada elo (se X, então Y, desde que Z) é plausível e bem articulado?
- **Base de evidências**: as hipóteses estão sustentadas por dados, literatura ou avaliações de experiências anteriores semelhantes?
- **Coerência externa:** a política está articulada a outras ações ou programas existentes? Há sobreposição ou sinergia? Ela se adequa bem ao contexto institucional, político, territorial e sociocultural?

A avaliação de desenho permite explorar criticamente a TdM proposta para a política. Essa análise parte do princípio de que toda política pública é uma aposta causal: seus formuladores acreditam que determinada sequência de ações — insumos, atividades e produtos — levará a resultados e impactos desejados.

No entanto, no mundo real, muitas políticas não contam com uma TdM explicitada. Nesses casos, a avaliação de desenho se volta à construção da TdM, preferencialmente de forma colaborativa com os principais atores envolvidos, como gestores, implementadores e beneficiários. Esse processo viabiliza a análise crítica e fortalece o alinhamento entre os componentes e os objetivos da política, de forma a permitir que perguntas avaliativas sejam definidas adequadamente.

Para examinar essa lógica, é necessário tornar explícitas as hipóteses que conectam os elementos da cadeia causal. Essas hipóteses representam suposições sobre como os efeitos desejados acontecerão — e precisam ser plausíveis, verificáveis e ajustadas ao contexto. Como reforçam autores como Rossi et al. (2004) e Vaessen (2017), uma boa avaliação de desenho precisa destrinchar essas hipóteses e testá-las à luz do conhecimento técnico e evidências empíricas.

Além das hipóteses, a avaliação de desenho deve mapear riscos de implementação, ou seja, fatores que podem comprometer a realização das etapas previstas da política, especialmente os elos iniciais da cadeia. São riscos operacionais, institucionais, logísticos ou contextuais que podem afetar negativamente a execução.

Neste guia, adotamos uma distinção didática entre:

- **Hipóteses de efetividade**: pressupostos que ligam os produtos entregues aos resultados esperados. Estão associadas à plausibilidade causal da política.
- **Riscos de implementação**: ameaças que afetam os elos iniciais da cadeia, da mobilização de insumos à entrega dos produtos.

Essa distinção ajuda a organizar o raciocínio avaliativo para estruturar a avaliação de desenho.

Exemplo prático 14— Avaliação de desenho do Programa de Reforço da Alimentação Escolar

Objetivo do Programa: Reduzir o absenteísmo escolar em regiões vulneráveis por meio da oferta diária de merenda de qualidade.

Hipótese central: *"Se os alunos receberem merenda de qualidade diariamente (incluindo merenda de produtos provenientes da agricultura familiar, produtos orgânicos, merenda livre de alimentos ultraprocessados), então frequentarão a escola com maior assiduidade."*

1. Coerência interna
A Teoria da Mudança apresentada foi considerada plausível, com elos bem estruturados entre oferta de merenda, aumento da frequência escolar, melhora nutricional e redução da evasão.

Análise: a lógica foi considerada consistente.
Recomendação: nenhuma mudança recomendada nessa dimensão.

2. Base de evidências

Análise: Revisões sistemáticas e outras evidências de políticas semelhantes indicam que a merenda escolar pode aumentar a frequência em contextos de insegurança alimentar (SNILSTVEIT et al., 2015; PMA, 2020), desde que a oferta seja regular, de qualidade e bem aceita pelos alunos.
Recomendação: garantir cardápios adequados ao contexto local, mecanismos de escuta dos alunos e monitoramento contínuo da entrega e da qualidade das refeições.

3. Coerência externa
A avaliação indicou que, em algumas regiões, causas estruturais como ausência de transporte escolar e necessidade de trabalho infantil eram fatores muito relevantes para a baixa frequência e evasão. Isso levantou dúvidas sobre a pertinência da alimentação escolar como resposta prioritária, frente a um problema de múltiplas causas. Além disso, o Programa Nacional de Alimentação Escolar (PNAE) já atendia parte das escolas, gerando risco de sobreposição.

Análise: o desenho original apresentava fragilidades na articulação com o contexto local e com outras políticas públicas.
Recomendação: redesenhar o programa para garantir complementaridade ao PNAE, priorizando escolas não cobertas e investindo na melhoria da qualidade da merenda. Também é recomendável reavaliar se a alimentação escolar, isoladamente, é a resposta mais estratégica para enfrentar o problema diagnosticado.

Conclusão: O programa mostrou-se promissor, mas exigia ajustes para garantir alinhamento com o contexto e com outras políticas públicas, além de assegurar sua efetividade frente às causas principais do problema.

Fonte: FGV CLEAR, elaboração própria com base em Snilstveit et al. (2015) e PMA (2020).

Aplicando na prática

Etapas para avaliar o desenho de uma política pública:

- (Re)construa a Teoria da Mudança e analise a coerência interna:

  Organize os elementos da política (insumos, atividades, produtos, resultados e impactos). Verifique se a cadeia causal faz sentido, se os elos estão logicamente articulados e se as hipóteses e os riscos estão bem definidos.

- Busque e mobilize evidências:

  Consulte evidências robustas, como revisões sistemáticas, avaliações anteriores de intervenções semelhantes e outras fontes confiáveis. Avalie se as hipóteses da política estão sustentadas por informações válidas.

- Analise a coerência externa:

  Avalie se a política está articulada a outras ações existentes. Identifique riscos de sobreposição, oportunidades de sinergia e verifique se a lógica da intervenção é adequada ao território, às capacidades institucionais e ao contexto sociocultural.

- Incorpore escuta qualificada:

  Envolva gestores, técnicos e beneficiários na análise (ou construção, se necessário) da Teoria da Mudança. Utilize oficinas, entrevistas ou grupos focais para validar hipóteses, identificar riscos e captar percepções que qualifiquem as análises anteriores.

- Ajuste o desenho e recomende melhorias:

  Com base em tudo que foi analisado, proponha ajustes na lógica da intervenção, na lista de riscos e hipóteses, nos mecanismos de implementação ou nos resultados esperados.

6 Plano de monitoramento e avaliação: usar a Teoria da Mudança para transformar estratégia em evidência[10]

A Teoria da Mudança (TdM) não é apenas uma ferramenta de planejamento. Ela também oferece a base lógica sobre a qual se constrói o sistema de monitoramento e avaliação (M&A) de uma política pública. Ao organizar de forma clara e estruturada o encadeamento entre insumos, atividades, produtos, resultados e impactos, a TdM permite identificar o que precisa ser acompanhado, quais indicadores devem ser utilizados e que perguntas avaliativas são mais relevantes para cada etapa.

Uma TdM bem construída:

- Explicita as hipóteses que conectam cada elo da intervenção;
- Identifica riscos que podem comprometer os efeitos pretendidos;
- Estabelece fundamentos lógicos para a seleção de indicadores;
- Orienta a formulação de perguntas que ajudam a entender se os objetivos foram alcançados, por que foram (ou não) alcançados e a que custo.

Por isso, o plano de M&A não deve ser elaborado ao final da política, como uma etapa isolada. Ele precisa ser pensado desde o início, integrado ao desenho da intervenção, e acompanhar todo o seu ciclo, da implementação à tomada de decisão.

Esse plano se desdobra em dois componentes distintos e complementares:

- **O plano de monitoramento**, voltado ao acompanhamento contínuo da execução e dos efeitos imediatos da política;
- **E o plano de avaliação**, que estrutura a produção de evidências sobre a qualidade do desenho, a fidelidade da implementação, os resultados e os impactos alcançados.

Ambos se apoiam na TdM para transformar a estratégia em evidência, e são orientados por uma lógica simples e poderosa: se sabemos onde queremos chegar e por quais caminhos, precisamos acompanhar se estamos no rumo certo — e aprender com o trajeto. Os subcapítulos a seguir detalham cada um desses componentes.

[10] A seguinte literatura fundamenta o conteúdo deste capítulo: Brasil (2018a), BetterEvaluation (2022), CIF (2024), Comissão Europeia (2024), GEF (2022), IJSN (2018c), Imas e Rist (2009), Kusek e Rist (2004), Patton (2008), Roberts e Khattri (2012), Rodriguez-Billella e Tapella (2024), Rossi, Lipsey e Freeman (2004), (OCDE, 2021, 2023).

6.1 Plano de Monitoramento: acompanhar para aprender, ajustar e entregar resultados

Se a TdM explicita os caminhos que a política espera percorrer, o plano de monitoramento é o instrumento que permite **acompanhar esses caminhos em tempo real**. Monitorar é acompanhar de forma sistemática, contínua e estruturada o que está sendo feito, o que está sendo entregue e que efeitos estão sendo gerados ao longo da implementação da política.

Mais do que um mecanismo de controle, o monitoramento é uma **ferramenta de gestão e de aprendizado**. Ele permite detectar desvios antes que eles se transformem em falhas, ajustar rumos de maneira oportuna, manter a coerência com os objetivos definidos e informar decisões durante toda a execução da política.

6.1.1 Como a Teoria da Mudança orienta o monitoramento?

Ao representar visualmente os elementos da intervenção e suas ligações causais, a TdM oferece uma **estrutura lógica para definir o que deve ser monitorado**. A partir dela, é possível identificar os principais **componentes a serem acompanhados** — insumos, atividades, produtos, resultados de curto e médio prazo — e estabelecer **indicadores correspondentes**.

Para cada elo da cadeia, o plano de monitoramento deve responder:

- O que precisa ser medido para saber se a política está no caminho certo?
- Como será feita a medição?
- Com que frequência?
- Quem é responsável por coletar e analisar os dados?
- Qual é a meta que se espera atingir?

Com base nessa estrutura, o plano de monitoramento deve incluir indicadores que reflitam cada etapa crítica da intervenção. Isso significa acompanhar, de forma articulada, os **insumos mobilizados** (como orçamento, disponibilidade de pessoal e materiais), as **atividades executada**s (como capacitações, atendimentos ou entregas), os **produtos gerados** (como refeições servidas, kits distribuídos, relatórios emitidos) e os **resultados imediatos esperados** (como aumento de frequência escolar ou melhora na adesão a serviços). Ao selecionar indicadores para cada um desses níveis, garante-se uma visão completa da execução da política e a possibilidade de agir com base em evidências ao longo do caminho.

Exemplo prático 15— Indicadores de produtos e resultados do Programa de Reforço da Alimentação Escolar

Retomando o nosso exemplo anterior, se a TdM prevê que a entrega de merenda escolar diária contribuirá para a frequência escolar, o monitoramento deve incluir:

- Indicador de produto: % de escolas que ofertam merenda 5 dias na semana;
- Indicador de resultado: variação da frequência escolar mensal;
- Ambos relacionados a metas definidas, contando com coleta periódica e responsáveis claros.

Fonte: FGV CLEAR, elaboração própria.

6.1.2 Componentes de um bom plano de monitoramento

Um plano de monitoramento deve contemplar cinco elementos básicos:

1. Indicadores relevantes

   Derivados da TdM, cobrindo todas as etapas críticas (insumos, atividades, produtos, resultados).

2. Metas e linha de base

   Toda meta deve ser comparada com um valor inicial (linha de base) e expressar o progresso esperado.

3. Periodicidade da coleta

   Indicadores de insumo e atividade geralmente são coletados com maior frequência; resultados podem demandar intervalos maiores.

4. Fontes de dados confiáveis

   Registros administrativos, sistemas próprios, surveys, observações ou dados secundários.

5. Responsabilidade institucional

   Quem coleta, consolida, analisa e decide com base nas informações.

Um bom plano de monitoramento não precisa ser complexo — precisa ser funcional, utilizado e orientado para a gestão. A recomendação é priorizar indicadores com relevância estratégica, clareza metodológica e custo viável de coleta.

Exemplo prático 16— Plano de Monitoramento Resumido para o Programa de Reforço da Alimentação Escolar

Indicador	Etapa da TdM	Linha de base	Meta	Frequência	Fonte de dados	Responsável
% de alunos com frequência ≥ 90%	Resultado	78%	90%	Mensal	Sistema de presença escolar	Secretaria de Educação
Nº de refeições servidas/dia	Produto	–	100% da matrícula	Diário	Relatório da merenda	Equipe escolar

Fonte: FGV CLEAR, elaboração própria.

Box 13 – Como escolher um indicador?

Ao desenhar indicadores para monitorar ou avaliar uma política pública, é essencial que os indicadores sigam os critérios conhecidos como SMART, um acrônimo amplamente utilizado no campo da gestão pública e avaliação:

- Específico: o indicador deve mensurar um aspecto claro e definido da política, sem ambiguidades;
- Mensurável: deve ser possível quantificá-lo ou descrevê-lo com precisão, com base em dados disponíveis ou coletáveis;
- Atingível: o nível de ambição deve ser realista, considerando os recursos, capacidades e restrições do contexto;
- Relevante: o indicador precisa estar diretamente relacionado aos objetivos e resultados da política;
- Temporal: deve conter uma definição clara de periodicidade e horizonte de tempo para coleta e análise.

Além desses critérios, recomenda-se envolver os gestores, técnicos e atores-chave na escolha dos indicadores. A participação desses profissionais garante que os indicadores selecionados sejam compreensíveis, úteis e efetivamente utilizados no processo decisório.

Fonte: Imas e Rist (2009); Roberts e Khattri (2012); Comissão Europeia (2024).

Aplicando na prática

Como construir um bom plano de monitoramento:

- Comece pela Teoria da Mudança: identifique os elementos que precisam ser acompanhados (insumos, atividades, produtos, resultados).
- Defina indicadores relevantes e viáveis — nem tudo pre-

cisa ser medido, mas tudo o que importa deve ser acompanhado.

- Estabeleça metas com base em uma linha de base confiável, e revise-as sempre que necessário.
- Planeje a periodicidade da coleta de acordo com a velocidade das mudanças esperadas.
- Deixe claro quem faz o quê: atribua responsabilidades pela coleta, análise e uso dos dados.
- Use os dados para ajustar a rota, e não apenas para prestar contas.

6.2 Plano de Avaliação: testar hipóteses, verificar riscos e aprender com os resultados

É também a partir da base lógica ilustrada pela Teoria da Mudança (TdM) que se constrói um **plano de avaliação robusto**, com perguntas claras, pertinentes e alinhadas à realidade da política. Esse plano organiza **o que se pretende avaliar, por que, quando, como e com base em quais dados**.

Exemplo prático 17— Avaliar é testar hipóteses da Teoria da Mudança

Cada seta da TdM representa uma hipótese causal — "se fizermos isso, então aquilo deve acontecer, desde que certas condições se mantenham". O papel da avaliação é testar essas hipóteses, com evidências e rigor, verificando se a Teoria do Programa se confirma na prática, onde ela precisa ser ajustada, e o que pode ser aprendido para o futuro.

Ilustrando a ideia acima a partir do nosso exemplo do programa de reforço à alimentação escolar, teríamos: "Se os alunos receberem merenda diária (atividade), então a frequência escolar aumentará (resultado), desde que a qualidade da alimentação seja aceitável e o transporte escolar funcione (condições)."

A TdM, portanto, não é apenas o ponto de partida para a formulação da política, mas também o alvo principal da avaliação.

Fonte: FGV CLEAR, elaboração própria.

6.2.1 A Teoria da Mudança como geradora de perguntas avaliativas

Para cada elo da cadeia lógica da política, dos insumos aos impactos, há uma ou mais **perguntas avaliativas potenciais**. Essas perguntas devem estar orientadas para:

- Verificar se a política foi implementada conforme o previsto;
- Confirmar se os resultados e impactos esperados foram alcançados;
- Investigar se as mudanças observadas são atribuíveis à política;
- Avaliar a qualidade da execução e a aderência às hipóteses formuladas;
- Compreender se a política valeu a pena, considerando custos e alternativas.

Box 14 – Etapas práticas para elaborar perguntas avaliativas a partir da Teoria da Mudança

1. Percorra a TdM, do impacto ao insumo, ou vice-versa;
2. Para cada elo, identifique a hipótese causal: "se X, então Y, desde que Z";
3. Pergunte-se:
    - Como saberei se essa transição ocorreu?
    - Que risco pode ter comprometido esse elo?
    - Que condição ou suposição precisa se confirmar?
4. Transforme essas reflexões em perguntas avaliativas claras e factíveis;
5. Classifique cada pergunta conforme o tipo de avaliação e fase da política;

Organize todas em um plano de avaliação, com indicadores, fontes e métodos.

Fonte: FGV CLEAR, elaboração própria.

6.2.2 Tipologia de perguntas avaliativas segundo a Teoria da Mudança

A Teoria da Mudança permite organizar o raciocínio causal da política pública em etapas encadeadas — insumos, atividades, produtos, resultados e impactos — que,

por sua vez, servem de base para a formulação de perguntas avaliativas ao longo do ciclo da política pública. Cada tipo de pergunta avaliativa se alinha a um momento específico do ciclo da política pública — da avaliação de desenho à avaliação de impacto e custo-benefício. A TdM funciona, assim, como uma ponte entre o raciocínio causal da política e o sistema de M&A que a acompanha. A seguir, a **Tabela 1** apresenta tipologia simplificada que relaciona perguntas típicas aos diferentes estágios da TdM e aos momentos do ciclo da política pública.

Tabela 1: Tipologia de perguntas avaliativas

Pergunta Avaliativa	**Fase ou tipo de avaliação**	**Objetivo da pergunta**
A entrega planejada ocorreu como previsto?	Implementação	Verificar a execução dos insumos e atividades conforme o desenho
O público-alvo foi atendido como esperado?	Resultados (curto prazo)	Confirmar se os produtos chegaram a quem deveriam
As metas previstas para os resultados foram alcançadas?	Resultados (intermediários)	Medir a efetividade dos resultados previstos
Os produtos entregues causaram as mudanças esperadas?	Impacto	Verificar se os resultados observados derivam da política
Os riscos se concretizaram?	Implementação	Investigar desafios que comprometeram a entrega
Houve problemas operacionais ou de contexto?	Implementação	Identificar obstáculos ou falhas que exigiram adaptações
A qualidade da implementação foi adequada?	Processo	Julgar o padrão da entrega dos produtos e serviços
Os impactos atribuíveis à política ocorreram?	Impacto	Analisar efeitos de longo prazo e sua relação com a intervenção
Os benefícios compensam os custos?	Custo-benefício / Custo-efetividade	Avaliar se os resultados justificam o investimento realizado
As atividades da TdM têm plausibilidade teórica para gerar os impactos esperados?	Avaliação de desenho	Julgar a consistência lógica e evidencial da cadeia causal

Fonte: FGV CLEAR, elaboração própria.

A Tabela 1 tem caráter ilustrativo: dentro de cada tipo de pergunta, ainda é possível explorar múltiplos ângulos, dimensões e recortes, utilizando diferentes tipos de metodologias, conforme o foco e os objetivos de cada avaliação.

Além da associação entre perguntas avaliativas e os estágios da TdM, é possível enriquecer ainda mais o plano de avaliação ao incorporar lentes analíticas transversais que aprofundam a compreensão sobre os efeitos da política. Entre essas lentes, destacam-se os **critérios da Organização para a Cooperação e Desenvolvimento Econômico (OCDE)**, amplamente utilizados como referência internacional para avaliar a qualidade e os resultados das intervenções públicas, e aborda-

gens específicas que consideram aspectos de **equidade, gênero, efeitos ambientais** (*Footprint*) e **mudança transformadora**. Essas dimensões complementam os tipos de avaliação ao oferecer diferentes ângulos de análise — seja para julgar a relevância, efetividade ou sustentabilidade da política, seja para entender como seus efeitos se distribuem entre grupos sociais ou contribuem para transformações sistêmicas (OCDE, 2021).

Os subcapítulos que seguem apresentam brevemente cada uma dessas abordagens, com orientações práticas sobre como integrá-las ao plano de avaliação.

6.2.3 Critérios analíticos da OCDE como lentes complementares

Ao estruturar um plano de avaliação, é comum organizar as perguntas segundo o momento da política pública — desenho, implementação, resultados, impacto, custo-benefício. Essa abordagem, ancorada na Teoria da Mudança (TdM), orienta a avaliação ao longo do tempo, testando hipóteses causais e verificando a lógica da intervenção.

No entanto, também é possível — e recomendável — incorporar **outras dimensões analíticas** que aprofundam o julgamento sobre a política pública. Entre essas, destacam-se os critérios de avaliação propostos pela OCDE, amplamente utilizados por governos, organismos multilaterais e avaliadores em diferentes contextos (OCDE, 2021).

Os seis critérios da OCDE apresentados no **Box 15** analisam diferentes níveis de uma política, conforme definição a seguir:

Box 15 – Critérios de avaliação da OCDE

A Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) criou os critérios de avaliação com o propósito de servirem como referência para a avaliação de projetos de desenvolvimento humanitários, programas e políticas públicas (OCDE, 2021). Criados pelo Comitê de Ajuda ao Desenvolvimento (DAC), os critérios auxiliam a verificar a qualidade da intervenção e se essa trouxe efeitos positivos ou negativos.

De forma conjunta e complementar, os seis critérios **(Figura 11)** refletem os Objetivos de Desenvolvimento Sustentável da Agenda 2030 e analisam diferentes níveis de uma política, conforme definição a seguir:

- **Relevância:** explicita a relação entre os objetivos, a implementação da política e seus beneficiários. Busca-se avaliar se o público-alvo entende a política como benéfica, importante e útil.

- **Efetividade:** busca responder se a intervenção está atingindo seus objetivos. Ou seja, a análise de efetividade verifica em que medida a política alcançou (ou tem potencial para alcançar) os resultados esperados.
- **Eficiência:** refere-se a como a busca avaliar se a intervenção está entregando seus resultados, tanto no uso dos recursos quanto no cumprimento dos prazos. Do ponto de vista econômico, analisa se os insumos — financeiros, humanos e naturais — estão sendo convertidos de forma custo-efetiva em produtos e impactos. No aspecto temporal, verifica se os resultados estão sendo entregues conforme o cronograma ou dentro de um prazo razoável, compatível com o contexto.
- **Impacto:** considera os efeitos da intervenção, sejam eles positivos ou negativos, intencionais ou não. Este critério explicita o potencial transformativo da intervenção ao identificar seus efeitos sociais, econômicos e até mesmo ambientais, dentro de uma perspectiva de longo prazo e de mudança estrutural (como mudanças no bem-estar populacional, direitos humanos e equidade de gênero).
- **Sustentabilidade:** identifica como os benefícios advindos da política perduram. Dentro do escopo deste critério, faz-se uma análise multidimensional dos sistemas necessários para a sustentação da política: examinam-se as capacidades financeiras, econômicas, sociais, ambientais e institucionais destes sistemas.
- **Coerência:** busca entender a compatibilidade da política analisada com as demais intervenções, que ocorrem em um país, setor ou instituição. Pode-se observar este critério de dois pontos de vista diferenteses: do ponto de vista do interno, busca-se compreender a sinergia e inter-relação da política avaliada com outras políticas geridas pela mesma instituição; do do ponto de vista externo, olha-se para a harmonização e coordenação com demais intervenções de outros atores, mas dentro do mesmo contexto.

Figura 11– Critérios de avaliação



Fonte: Adaptado de OCDE (2021).

6.2.4 Como integrar critérios da OCDE nos diferentes tipos de avaliação?

Os critérios da OCDE não substituem os diferentes tipos de avaliação existentes — como as avaliações de desenho, implementação, resultados ou impacto. Ao contrário, eles funcionam como filtros analíticos complementares, que podem ser utilizados dentro de qualquer tipo de avaliação, para orientar o foco das perguntas, qualificar os julgamentos e aprofundar a análise.

Por exemplo:

- Em uma avaliação de desenho, os critérios de relevância e coerência ajudam a analisar se a política está bem alinhada ao problema diagnosticado e se articula com outras iniciativas existentes.
- Em uma avaliação de implementação, o critério de eficiência pode orientar a análise sobre o uso de recursos e prazos de execução.
- Em uma avaliação de resultados, o foco pode recair sobre a efetividade, analisando se os resultados esperados foram atingidos.
- Em uma avaliação de impacto, é possível avaliar não apenas a dimensão de impacto, mas também analisar aspectos de sustentabilidade.

- Em avaliações de desenho, o critério de coerência permite observar sinergias ou sobreposições com outras políticas no mesmo contexto.

A OCDE (2021) recomenda dois princípios fundamentais para o uso adequado de seus critérios em avaliações:

1. Aplicação cuidadosa e contextualizada

   Os critérios devem ser interpretados à luz do propósito da avaliação, do momento em que ela ocorre e do contexto institucional e territorial. Devem ser adaptados à realidade da política avaliada, considerando disponibilidade de dados, recursos e objetivos.

2. Seleção crítica e estratégica dos critérios

   Nem todas as avaliações exigem a aplicação de todos os critérios. A escolha deve refletir os objetivos do estudo, as prioridades dos usuários da avaliação e a viabilidade de análise. O uso dos critérios deve ser reflexivo, e não mecânico.

O **Box 16** traz exemplos de uso dos critérios da OCDE em diferentes tipos de avaliação.

Box 16 – Exemplos de uso dos critérios da OCDE em diferentes tipos de avaliação

Tipo de Avaliação	Critério da OCDE	Exemplo de pergunta avaliativa
Avaliação de Desenho	Relevância / Coerência	A política está alinhada ao problema e articulada com outras ações existentes?
Avaliação de Implementação	Eficiência	Os recursos foram bem utilizados, e os prazos foram respeitados?
Avaliação de Implementação	Coerência	A execução da política dialoga com outras ações governamentais no mesmo contexto?
Avaliação de Resultados	Efetividade	Os produtos entregues geraram os resultados esperados?
Avaliação de Impacto	Sustentabilidade / Impacto	Os efeitos da política se mantêm no tempo? Quais transformações sociais ocorreram?
Avaliação de Custo-Benefício	Eficiência / Impacto	O investimento compensa os benefícios gerados?

Fonte: FGV CLEAR, elaboração própria.

6.2.5 Outras abordagens para enriquecer o plano de avaliação: Avaliação com foco em equidade e gênero

A avaliação com foco em equidade e gênero parte do reconhecimento de que políticas públicas não afetam todas as pessoas da mesma forma. Diferentes grupos sociais — como mulheres, pessoas não binárias, negras, indígenas, com deficiência, entre outros — vivenciam **desigualdades históricas e estruturais** que moldam sua capacidade de acessar, permanecer e se beneficiar de políticas e programas (OCDE, 2021).

Essa abordagem propõe que o plano de avaliação incorpore, de forma sistemática, questões relacionadas à justiça social, à não discriminação e à inclusão, desde a formulação das perguntas até a análise dos resultados e a elaboração das recomendações. Trata-se de garantir que todos sejam considerados, e que as desigualdades existentes sejam explicitamente analisadas e enfrentadas, com o objetivo de promover transformações efetivas nas políticas públicas.

Box 17 – Como incorporar uma abordagem equitativa e de gênero no plano de avaliação?

Incluir a lente de equidade e gênero no plano de avaliação significa, antes de tudo, formular perguntas avaliativas específicas sobre quem se beneficia da política, quem pode estar sendo ignorado e quais são os efeitos diferenciados para distintos grupos sociais. Além disso, é importante:

- Desagregar dados por sexo, raça, deficiência, idade ou território, sempre que possível;
- Escolher métodos e fontes de dados que permitam capturar desigualdades, como grupos focais específicos, entrevistas com lideranças comunitárias ou mapeamento participativo;
- Incluir grupos sub-representados no processo avaliativo, seja na definição das perguntas, na coleta de dados ou na validação dos achados;
- E, por fim, analisar as barreiras de acesso, permanência e apropriação dos resultados da política, com recomendações concretas para superá-las.

Fonte: FGV CLEAR, elaboração própria.

6.2.6 Outras abordagens para enriquecer o plano de avaliação: Avaliação participativa

A avaliação participativa é uma abordagem que propõe o **engajamento ativo das partes interessadas** — gestores, técnicos, beneficiários, lideranças comunitárias e organizações da sociedade civil — em diferentes fases do processo avaliativo. Seu princípio central é que os sujeitos envolvidos na política pública **possuem conhecimentos, experiências e perspectivas valiosas** que devem ser consideradas na definição do que será avaliado, de como será avaliado e de como os achados serão interpretados e utilizados (RODRIGUEZ-BILLELLA & TAPELLA, 2024).

Diferente da avaliação convencional, que frequentemente é conduzida por avaliadores externos com pouca interação com os participantes da política, a abordagem participativa entende que a avaliação deve ser um espaço de **diálogo, escuta e construção coletiva**. O grau de participação pode variar desde consultas pontuais (por meio de entrevistas, oficinas ou questionários) até modelos mais profundos de cocriação, em que os próprios participantes contribuem na formulação das perguntas, na coleta e interpretação dos dados e até na formulação de recomendações (RODRIGUEZ-BILLELLA & TAPELLA, 2024).

Box 18 – Como incorporar a avaliação participativa no plano de avaliação

Para adotar essa abordagem, o plano de avaliação pode prever:

- Atores sociais como coplanejadores da avaliação, e não apenas fontes de informação;
- Etapas de escuta ativa, como oficinas participativas para definir perguntas avaliativas ou validar achados;
- Definição compartilhada dos usos da avaliação, com atenção aos objetivos de aprendizado, responsabilização e transformação.

Fonte: FGV CLEAR, elaboração própria.

6.2.7 Outras abordagens para enriquecer o plano de avaliação: *Footprint Evaluation* e justiça climática como lentes transversais

A crise climática desafia não apenas políticas ambientais, mas toda a ação pública. Programas de habitação, transporte, infraestrutura, educação ou agricultura têm efeitos diretos ou indiretos sobre o meio ambiente e são afetados pelas mudanças no clima. Diante disso, cresce a demanda por avaliações que incorporem a dimensão climática como uma lente transversal, independentemente do setor de atuação

da política (BETTEREVALUATION, 2022).

É nesse contexto que surge a abordagem conhecida como *Footprint Evaluation*. O projeto *Footprint Evaluation* é uma iniciativa de cocriação e pesquisa coordenada pela plataforma BetterEvaluation, com apoio da Global Evaluation Initiative, voltada a integrar a sustentabilidade ambiental como lente transversal nas avaliações de políticas e programas (BETTEREVALUATION, 2022).

Esta abordagem propõe avaliar a pegada ecológica e climática de qualquer política pública, mesmo que seu objetivo principal não seja ambiental. Trata-se de uma mudança de perspectiva: em vez de avaliar apenas os efeitos diretos e intencionais de uma política, a Footprint Evaluation propõe olhar também para os **impactos ambientais indiretos, colaterais e de longo prazo**, sejam desejados ou não (BETTEREVALUATION, 2022).

A abordagem também incorpora princípios da **justiça climática**, reconhecendo que os riscos e benefícios ambientais não são igualmente distribuídos na sociedade. Populações historicamente marginalizadas, como comunidades periféricas, indígenas, ribeirinhas ou trabalhadores informais, tendem a enfrentar maior exposição a riscos climáticos e menor acesso às oportunidades geradas pelas políticas públicas. Avaliar os efeitos distributivos das intervenções é, portanto, um componente essencial dessa lente.

Box 19 – Como incorporar essa abordagem no plano de avaliação

A *Footprint Evaluation* pode ser utilizada como uma lente transversal em qualquer plano de avaliação, especialmente nas seguintes etapas:

- Formulação das perguntas avaliativas:
  - A política em questão contribui para aumentar ou reduzir emissões?
  - Há impacto sobre o uso do solo, dos recursos naturais ou da biodiversidade?
  - Os benefícios e riscos ambientais são igualmente distribuídos entre os grupos sociais?
- Escolha dos critérios analíticos:
  - Sustentabilidade e coerência ganham centralidade, agora com enfoque climático;
  - Justiça climática passa a orientar a análise distributiva.
- Definição de indicadores e fontes de dados:
  - Emissões evitadas, uso energético, pegada hídrica, capacidade adaptativa local, entre outros;
  - Dados geoespaciais, sensores, registros ambientais, entrevistas com comunidades afetadas.

Mesmo que a política avaliada não tenha um enfoque ambiental, pergunte: "Essa intervenção tem efeitos sobre o clima ou sobre grupos vulneráveis ao clima?" Se sim, a lente da Footprint Evaluation pode trazer achados valiosos, inclusive para decisões futuras de financiamento e escala.

Fonte: BetterEvaluation (2022).

6.2.8 Outras abordagens para enriquecer o plano de avaliação: Avaliação de mudanças transformadoras

A noção de *transformational change* (mudança transformacional, se traduzido literalmente para a língua portuguesa) tem sido desenvolvida por instituições como o Fundo Global para o Meio Ambiente (*Global Environmental Facility* – ou GEF – em Inglês) e os Fundos de Investimento Climático (*Climate Investment Funds* – ou CIF – em Inglês) para avaliar políticas e iniciativas que promovem mudanças relevantes, profundas, estruturais e sustentáveis, por exemplo, reformas institucionais, transformações de mercado ou mudanças sistêmicas em áreas como energia, biodiversidade e clima (GEF, 2022; CIF, 2024). Embora tenha origem em agendas ambientais, essa abordagem pode ser útil para qualquer política que busque **romper padrões históricos, criar capacidades ou promover efeitos sustentáveis e de longo prazo com potencial de disseminação**. Avaliar esse tipo de mudança ajuda a entender se a política realmente alterou o funcionamento do sistema — e não apenas entregou produtos.

Box 20 – Como incorporar a abordagem de mudanças transformacionais no plano de avaliação

A avaliação de mudanças transformacionais pode ser incorporada como uma lente complementar para identificar efeitos sistêmicos, sustentáveis e com potencial de escala. Ela ajuda a entender se a intervenção alterou não apenas indicadores imediatos, mas também sistemas, práticas institucionais ou relações de poder.

Para isso, é importante observar sinais e formular perguntas avaliativas que explorem se houve mudanças institucionais, disseminação ou apropriação dos efeitos, e se os impactos tendem a ser duradouros. A análise pode se apoiar em critérios como sustentabilidade, coerência e transformações estruturais que se mantêm no tempo, buscando evidências capazes de identificar essas mudanças — como entrevistas com atores estratégicos, análise documental e estudos de caso qualitativos.

Mesmo em políticas sociais ou territoriais, vale perguntar: a intervenção iniciou alguma mudança estrutural? Se sim, a lente da mudança transformacional pode ampliar a compreensão do impacto e oferecer subsídios para mantê-las no tempo e levá-las a outros contextos.

Fonte: Elaboração própria com base em CIF (2024) e GEF (2022).

6.2.9 Estruturando o plano de avaliação

Com as perguntas definidas, o plano de avaliação organiza todas as informações necessárias para que possam ser respondidas com qualidade, dentro dos prazos e com os dados adequados. O exemplo prático apresentado abaixo desenvolve um plano de avaliação para o Programa de Reforço à Alimentação Escolar que vem ilustrando nossos exemplos, a partir do que foi discutido ao longo desta seção.

Exemplo prático 18— Programa de Reforço da Alimentação Escolar

Pergunta Avaliativa	Tipo de Avaliação	Lentes Adicionais	Fonte de Dados	Responsável
A política está bem desenhada para alcançar o impacto esperado?	Avaliação de desenho	Relevância, Coerência, Gênero, Equidade	TdM, revisões sistemáticas, entrevistas com especialistas	Coordenação central da política
A política foi desenhada com potencial para gerar mudanças estruturais no sistema educacional?	Avaliação de desenho	Relevância, Sustentabilidade, Transformational change	Entrevistas com gestores, análise documental	Equipe de gestão + avaliadores externos
Os insumos e atividades foram executados como planejado?	Avaliação de implementação	Eficiência, Coerência, Footprint, Participativa	Relatórios mensais, dados administrativos, supervisões de campo	Equipe de implementação local
A alimentação foi oferecida com regularidade e qualidade?	Avaliação de processo	Efetividade, Sustentabilidade, Gênero, Equidade	Monitoramento escolar, entrevistas com usuários	Coordenação local do programa
A política é custo-efetiva?	Análise de custo-benefício	Eficiência, Impacto	Dados administrativos e registros escolares	Secretaria de Educação e Secretaria de Finanças
Houve variações nos efeitos por sexo, raça ou território?	Avaliação de resultados ou impacto	Gênero, Equidade, Justiça social	Registros escolares, entrevistas com grupos específicos	Monitoramento com especialistas externos
A política gerou algum efeito ambiental positivo ou negativo?	Avaliação de resultados ou impacto	Sustentabilidade, Footprint, Justiça climática	Dados administrativos, observações em campo	Coordenação local + equipe ambiental
A política aumentou a frequência escolar?	Avaliação de impacto	Sustentabilidade, Transformational change, Gênero, Equidade	Registros escolares, dados administrativos	Secretaria de Educação

Fonte: FGV CLEAR, elaboração própria.

Aplicando na prática

Como estruturar um bom plano de avaliação

- Use a Teoria da Mudança como ponto de partida: cada elo da cadeia causal pode gerar uma ou mais perguntas avaliativas.
- Priorize as perguntas que ajudam a tomar decisões estratégicas, melhorar a política ou justificar sua continuidade.
- Escolha os tipos de avaliação mais adequados ao momento e aos objetivos: desenho, implementação, resultados, impacto, custo-benefício.
- Defina indicadores, fontes de dados e métodos compatíveis com cada pergunta.
- Considere lentes analíticas transversais, como equidade, Footprint, sustentabilidade ou coerência com outras políticas.
- Planeje o uso dos resultados desde o início: quem vai usar, para quê, e como os achados serão comunicados.

7 Conclusão

A análise ex ante é uma etapa essencial para fortalecer a qualidade e maximizar o potencial de sucesso das políticas públicas, desde a sua concepção. Ao incorporar o uso sistemático de evidências no diagnóstico dos problemas, na definição dos públicos-alvo, na escolha de alternativas e no desenho da Teoria da Mudança, gestores e equipes técnicas ampliam a capacidade de formular políticas mais coerentes, realistas e transformadoras. O planejamento informado por evidências permite antecipar riscos, testar hipóteses, estruturar estratégias sólidas e estruturar sistemas de monitoramento e avaliação capazes de apoiar a implementação e o aprimoramento contínuo das ações. Mais do que uma etapa preliminar, a análise ex ante torna factível e intencional o sucesso futuro da política, prezando pela boa gestão dos recursos públicos e pela promoção de resultados que impactem positivamente a vida das pessoas. Nas próximas seções, discutiremos as estratégias de análise ex post, completando o percurso de uma gestão pública orientada para o aprendizado e para a excelência na entrega de valor à sociedade.

PARTE III

Análise *ex post*: analisar, aprender e aprimorar

Parte III – Análise ex post: analisar, aprender e aprimorar

A análise ex post é realizada durante ou após a implementação de uma política pública, com o objetivo de acompanhar sua execução, examinar o que foi feito, verificar quais resultados foram alcançados e identificar caminhos para o aprimoramento das ações em curso. Essa etapa envolve métodos e abordagens que permitem verificar se a política pública foi implementada conforme o planejado, se alcançou o público-alvo com qualidade e se gerou os resultados esperados para a sociedade.

Nesta Parte III, apresentamos seis modalidades de análise ex post, cada uma com seus objetivos específicos:

1. Monitoramento da implementação
2. Avaliação de implementação e processos
3. Avaliação de resultados
4. Avaliação de impacto
5. Avaliação de custo-benefício e custo-efetividade
6. Avaliação executiva

Essas abordagens podem ser aplicadas de forma complementar e articulada, de acordo com a maturidade da política, as perguntas avaliativas formuladas e o tipo de decisão que se deseja embasar. Enquanto o monitoramento e a avaliação de implementação e de processos ajudam a entender como a política foi colocada em prática, as avaliações de resultados e de impacto permitem mensurar os efeitos gerados. Já as análises de custo-benefício e de custo-efetividade examinam a relação entre os recursos utilizados e os benefícios sociais gerados, enquanto a avaliação executiva oferece uma visão integrada de forma ágil, especialmente útil em contextos de tomada de decisão com prazos reduzidos.

Conforme apresentado na Parte II, a avaliação de desenho permite uma análise crítica da concepção da política. Embora seja especialmente útil no momento ex ante, quando ainda é possível aprimorar o desenho antes da implementação, ela também pode ser realizada ex post, para investigar se eventuais falhas nos resultados decorreram de problemas estruturais no desenho. Em ambos os casos, o foco está na verificação da plausibilidade dos mecanismos causais, da fundamentação dos pressupostos e da coerência do desenho em relação ao contexto. Por isso, ainda que não esteja listada entre as modalidades centrais desta parte, a avaliação de desenho é uma ferramenta relevante a ser considerada no conjunto das estratégias ex post — e frequentemente realizada dentro do escopo da avaliação executiva.

Ao longo desta Parte III, discutiremos como essas diferentes estratégias avaliativas podem contribuir para o aperfeiçoamento de políticas em curso, fortalecendo a gestão pública baseada em evidências e apoiando decisões sobre sua continuidade, expansão ou reformulação.

Este guia oferece uma visão geral e prática das principais abordagens e métodos de avaliação ex post de políticas públicas e programas sociais. Seu intuito é facilitar a compreensão e orientar a escolha das técnicas mais adequadas ao contexto específico de cada política ou programa. Para aprofundamentos adicionais, recomenda-se consultar publicações especializadas da série Avaliação na Prática do FGV CLEAR e as referências indicadas ao final deste documento.

1 Monitoramento: colocando o plano em prática[11]

Como discutido na Parte II deste guia, o plano de monitoramento é construído com base na Teoria da Mudança (TdM) da política e define o que será acompanhado, com que frequência, por quem e com que finalidade. Ele transforma as etapas causais da intervenção em perguntas operacionais, oferecendo uma base estruturada para a gestão orientada por evidências.

Nesta Parte III, o foco passa a ser a **aplicação concreta** desse plano: como transformá-lo em uma rotina prática de acompanhamento, aprendizado e melhoria contínua durante a fase de implementação.

Ao longo da execução de uma política pública, o monitoramento permite verificar se os insumos estão sendo mobilizados conforme o previsto, se as atividades estão sendo realizadas no prazo desejado, se os produtos estão sendo entregues com regularidade e qualidade, e se os primeiros efeitos esperados começam a se manifestar. Mais do que coletar dados, monitorar é usar informação em tempo real para orientar decisões com maior agilidade, eficiência e legitimidade.

Um bom plano de monitoramento não precisa ser complexo — mas precisa ser funcional, utilizável e orientado à gestão. Para isso, conforme apresentado na Parte II, ele deve contemplar cinco elementos essenciais:

- **Indicadores relevantes**, derivados da TdM e distribuídos ao longo da cadeia lógica da política (insumos, atividades, produtos e resultados[12]);
- **Metas e linha de base**, que permitam comparar os indicadores com valores de referência e acompanhar o progresso ao longo do tempo;
- **Periodicidade adequada de coleta**, considerando a natureza de cada elemento observado e o ritmo de cada informação;
- **Fontes de dados confiáveis**, como registros administrativos, sistemas próprios ou instrumentos de pesquisa;

[11] Este capítulo se baseia, como um todo, nos trabalhos de Kusek e Rist (2004), Bamberger e Angeli (2009), Bamberger e Hewitt (1986), Patton (2008), Bamberger, Rugh e Mabry (2012), OCDE (2023), Rossi, Lipsey e Freeman (2004), Roberts e Khattri (2012), Imas e Rist (2009), IJSN (2018c).

[12] Indicadores de impacto não costumam ser úteis para a gestão cotidiana porque, em geral, refletem mudanças que levam tempo para acontecer e podem ser influenciadas por muitos fatores externos à política. Além disso, embora possam ser medidos com dados disponíveis, sua interpretação exige cautela, pois não é simples atribuir mudanças observadas diretamente à intervenção. Por isso, eles não são o foco do monitoramento, que deve priorizar o acompanhamento de insumos, atividades, produtos e resultados imediatos — ou seja, aquilo que está sob o controle direto da gestão.

- **Responsabilidades institucionais definidas**, que indiquem claramente quem coleta, quem consolida, quem analisa e quem usa as informações.

Esses cinco elementos estruturam o monitoramento como uma ferramenta estratégica. Um plano bem executado deve produzir dados úteis para identificar gargalos, corrigir desvios, reforçar boas práticas e orientar decisões corretivas com rapidez.

Por outro lado, um plano de monitoramento composto por muitos indicadores pode se tornar muito complexo, requerendo assim recursos demasiados para a coleta de todos os dados necessários. Isso torna necessária a priorização de indicadores. Roberts e Khattri (2012, p.38) propõem as seguintes perguntas norteadoras para orientar a escolha dos indicadores[13]:

- O indicador é necessário para monitorar o caminho até o objetivo desejado?
- A introdução do indicador pode levar à sobrecarga de quem estiver fornecendo ou coletando os dados?
- Qual é a contribuição deste indicador para o monitoramento e a avaliação da intervenção?

Além disso, é fundamental que os dados gerados sejam, de fato, utilizados. Dados de baixa qualidade, desatualizados ou ignorados na gestão não apenas perdem seu valor, mas também podem distorcer diagnósticos e comprometer decisões. Afinal, **monitorar com qualidade é monitorar com propósito — e com uso**.

1.1 O que o monitoramento permite observar?

Quando utilizado de forma sistemática e orientado por uma TdM bem formulada, o monitoramento permite **acompanhar com clareza o progresso da política em direção aos seus objetivos**, fornecendo respostas relevantes para a gestão em tempo real. Ao observar as principais etapas da cadeia lógica da política — insumos, atividades, produtos e primeiros resultados — os dados gerados possibilitam avaliar se a implementação está ocorrendo como o planejado e se há sinais iniciais de que os efeitos desejados começarão a emergir.

As principais perguntas que o monitoramento busca responder incluem:

- Os insumos foram alocados conforme o previsto?

[13] Perguntas originais em Inglês: "Is this indicator absolutely necessary to measure whether progress toward the strategic objective is being achieved? Will it create additional burdens on the respondents or on the staff collecting the data? How will this indicator help with monitoring, management, and evaluation?"

- As atividades estão sendo realizadas no prazo, escopo e frequência esperados?
- Os produtos (entregas diretas) estão sendo ofertados com a regularidade prevista, qualidade e em volume compatível com as metas?
- O público-alvo está sendo efetivamente alcançado? Há subgrupos com menor cobertura?
- Existem variações relevantes entre regiões, unidades de implementação ou perfis de beneficiários?
- Há gargalos que exigem ações corretivas imediatas?

Essas questões ajudam a antecipar desvios, alinhar ações com metas operacionais e informar a priorização de esforços. Responder a elas **fortalece a capacidade de a política pública gerar resultados concretos**, de forma mais equitativa, eficiente e ajustada ao seu contexto real.

Vale destacar que, conforme a literatura especializada em monitoramento e avaliação (M&A), **monitorar as atividades realizadas e os produtos entregues é tão relevante — e, muitas vezes, mais decisivo para a gestão imediata — do que acompanhar resultados finais**. Enquanto os resultados tendem a se manifestar ao longo do tempo, os produtos são os entregáveis que estão diretamente sob o controle da política e que condicionam os próximos passos da cadeia de causalidade.

Por exemplo, em um programa de reforço da alimentação escolar, saber se a merenda chegou a todas as escolas nos dias planejados (produto) é mais útil para a gestão, do ponto de vista operacional, do que esperar por dados semestrais de melhoria na aprendizagem (resultado). **Sem a entrega efetiva do produto, os resultados não virão**. Por isso, o monitoramento de produtos é o principal termômetro da execução no curto prazo e permite ajustes rápidos antes que falhas comprometam os impactos futuros.

Um bom sistema de monitoramento observa o que está sendo feito, o que está sendo entregue e os efeitos iniciais dessas ações. Essa visão integrada apoia decisões mais ágeis e bem fundamentadas.

Box 21 – Tipos de Indicadores para Monitoramento

Indicadores são ferramentas essenciais para transformar a Teoria da Mudança em evidência útil. Um bom plano de monitoramento inclui diferentes tipos de indicadores, organizados de acordo com os elementos da cadeia causal da política, desde a mobilização de recursos até os primeiros efeitos gerados:

Tipo de Indicador	O que mede	Exemplo
Insumo	Recursos mobilizados	Valor do orçamento executado
Atividade	Ações realizadas	Nº de formações para servidores
Produto (output)	Entregas diretas ao público	Nº de refeições servidas
Resultado (outcome)	Efeitos de curto ou médio prazo	Variação na frequência escolar

Além da posição na cadeia de resultados, os indicadores também podem ser classificados por outras dimensões relevantes para o monitoramento (OCDE, 2023; Rossi, Lipsey & Freeman, 2004):

- Indicadores quantitativos: expressam fenômenos em números (valores absolutos, percentuais, médias), como número de alunos beneficiados ou taxa de abandono escolar. São úteis para medir cobertura, volume e evolução ao longo do tempo.

- Indicadores qualitativos medem percepções, opiniões e aspectos subjetivos da realidade, como o grau de satisfação dos usuários com os serviços recebidos ou o nível de confiança no atendimento. Esses indicadores são geralmente coletados por meio de entrevistas, observações ou questionários com escalas de resposta — por exemplo, o nível de satisfação pode ser medido em uma escala de 1 a 5, variando de "muito insatisfeito" a "muito satisfeito".

- Indicadores simples: também conhecidos como indicadores unidimensionais, medem diretamente uma única variável ou dimensão observável, com cálculo e interpretação diretos. Exemplo: "número de refeições servidas por dia".

- Indicadores sintéticos ou compostos: também conhecidos como indicadores multidimensionais, combinam diferentes variáveis em uma única medida agregada, permitindo representar fenômenos complexos de forma simplificada. Um exemplo é o Índice de Desenvolvimento da Educação Básica (IDEB), que reúne em um único número dimensões como o desempenho dos alunos em provas padronizadas e a taxa de aprovação escolar. Esses indicadores exigem maior sofisticação metodológica, mas oferecem uma visão integrada e facilitam a comunicação e o acompanhamento de resultados.

Não existe um tipo de indicador melhor do que outro. A escolha mais adequada depende do objetivo do monitoramento, da qualidade dos dados disponíveis e do tipo de decisão que se deseja informar. Em geral, recomenda-se combinar diferentes tipos de indicadores — quantitativos e qualitativos, simples e compostos — para possibilitar uma visão mais completa, realista e útil para a gestão.

Todos os indicadores devem ser construídos com clareza metodológica e atender aos critérios SMART (específicos, mensuráveis, atingíveis, relevantes e temporais), como discutido na Parte II deste guia.

Fonte: FGV CLEAR, elaboração própria com base em OCDE (2023), Rossi, Lipsey e Freeman (2004), Roberts e Khattri (2012), Imas e Rist (2009).

1.2 Monitoramento orientado ao uso da informação

Para que o monitoramento seja efetivo, é fundamental que os dados produzidos sejam utilizados de forma estratégica na tomada de decisão. Isso exige clareza sobre quem coleta as informações, quem as analisa, quem as utiliza e como elas circulam entre os atores envolvidos. Mais do que gerar dados, é preciso garantir que eles alimentem rotinas de gestão, planejamento e prestação de contas.

Um sistema de M&A eficaz deve ser construído com base em demandas claras por informação e em mecanismos institucionais que assegurem o uso ativo dos dados. Ou seja, não basta medir: é essencial integrar a informação aos ciclos decisórios da política pública.

Para isso, é importante prever:

- Reuniões periódicas baseadas em evidências;
- Alertas para metas não atingidas;
- Painéis que facilitem a visualização e interpretação de indicadores;
- Planos de ação com responsabilidades definidas e acompanhamento de sua execução;
- Comunicação fluida entre áreas técnicas, gestores locais e as equipes responsáveis pela implementação.

Box 22 – Transformando dados em decisões efetivas

Coletar bons dados é apenas o primeiro passo. Para que eles realmente façam diferença, é essencial que cheguem às mãos certas, no momento certo e em formatos que possibilitem a ação. Mais do que produzir informação, é preciso criar condições para seu uso efetivo.

Isso envolve distribuir responsabilidades, definir protocolos claros de resposta e estabelecer um ambiente organizacional propício à tomada de decisão informada. Mesmo arranjos simples — como comissões técnicas com atribuições explícitas e acesso direto aos dados — podem gerar impactos significativos na qualidade da implementação.

O uso da informação só ocorre de forma consistente quando há liderança institucional comprometida com a aprendizagem e estruturas que garantam o fluxo de dados até os tomadores de decisão.

Fonte: FGV CLEAR, elaboração própria com base em Bamberger e Angeli (2009).

Exemplo prático 19— Monitoramento no Programa Bolsa Família

O Programa Bolsa Família é um dos exemplos mais consolidados de uso estratégico do monitoramento no Brasil. Desde sua criação, o governo federal disponibiliza painéis públicos de acompanhamento com indicadores atualizados sobre a gestão, cobertura e cumprimento das condicionalidades do programa.

Esses painéis permitem:

- Acompanhar o número de famílias beneficiadas, por município ou estado;
- Verificar o cumprimento das condicionalidades de saúde e educação (como frequência escolar mínima e vacinação);
- Monitorar indicadores de gestão descentralizada (como o IGD – Índice de Gestão Descentralizada), que avaliam a qualidade da execução local.

Essas informações são utilizadas por gestores municipais, órgãos de controle e pela própria sociedade, promovendo transparência e prestação de contas aos cidadãos, além de aprendizado e aprimoramento contínuo da política. A apresentação dos dados, com atualização periódica, série histórica e filtros por território, torna o painel uma ferramenta exemplar de transparência ativa e gestão baseada em evidências. É um caso reconhecido internacionalmente como referência de sistema de monitoramento eficaz em políticas sociais.

Fonte: FGV CLEAR, elaboração própria com base em Brasil (s.d.).

1.3 Monitoramento e transparência

Além de apoiar a gestão, o monitoramento pode — e deve — ser um instrumento de **transparência ativa e engajamento social**. Quando os dados de desempenho de uma política são divulgados de forma clara, acessível e atualizada, criam-se condições para que cidadãos, conselhos, a imprensa e organizações da sociedade civil acompanhem a execução, façam perguntas qualificadas e participem do aprimoramento das políticas.

A transparência é uma das condições para o uso estratégico das evidências. Ela reforça a responsabilização dos gestores, favorece a cooperação entre níveis de governo e estimula a participação informada dos cidadãos.

Experiências bem-sucedidas de monitoramento público, como a apresentada no exemplo prático 19 sobre o Programa Bolsa Família, atestam que divulgar dados com regularidade, em linguagem acessível e com recortes territoriais claros pode fortalecer a legitimidade das políticas públicas, facilitar o controle social e incen-

tivar melhorias na execução.

Ao tornar seus dados visíveis e compreensíveis, uma política pública envia um sinal claro: está aberta ao escrutínio, confia nos resultados que produz e **valoriza o diálogo com a sociedade como parte da sua estratégia de aperfeiçoamento contínuo**.

Aplicando na prática

Um plano de monitoramento só gera valor quando se transforma em prática cotidiana de gestão.

- Priorize indicadores que realmente importam e possam orientar decisões. Dados que não são utilizados acabam se tornando um custo, e não uma evidência útil.
- Estabeleça rotinas de análise e resposta. Painéis e alertas informam, mas decisões mudam cenários.
- Use o monitoramento para comunicar, ajustar e aprender — com a equipe, com parceiros e com a sociedade.
- O melhor momento para corrigir um problema é antes que ele comprometa os resultados.

2 Avaliação de Implementação e Processos de Políticas Públicas[14]

A avaliação de implementação e a avaliação de processos são estudos que analisam com profundidade como uma política pública ou programa está sendo executado na prática, indo além das informações oferecidas pelo monitoramento. Enquanto o monitoramento acompanha continuamente a execução com base em indicadores previamente definidos, essas avaliações investigam, de forma estruturada, as dinâmicas internas da implementação — a chamada "caixa-preta" da intervenção — buscando entender por que a política está (ou não está) operando conforme o previsto. Elas examinam se as atividades estão ocorrendo como planejado, se os serviços são entregues com qualidade, se o público-alvo está sendo efetivamente atendido e quais fatores contextuais, operacionais ou institucionais influenciam esses processos.

Esse tipo de avaliação permite verificar se os principais componentes da intervenção estão sendo implementados como previsto. Isso inclui examinar se as atividades estão ocorrendo nos prazos e formatos definidos; se os produtos e serviços são entregues com a qualidade esperada; se o público-alvo priorizado está sendo efetivamente atendido; se a equipe responsável possui capacitação e estrutura adequadas; e se os recursos financeiros e materiais são suficientes e estão sendo utilizados de forma eficaz.

Embora frequentemente utilizados de forma intercambiável, os termos **avaliação de implementação** e **avaliação de processos** apresentam ênfases distintas e complementares.

A **avaliação de implementação** concentra-se na fidelidade ao desenho da política ou programa. Seu objetivo é verificar se a execução ocorreu conforme o planejado, especialmente no que se refere às atividades realizadas, à cobertura alcançada e à alocação de recursos. Perguntas típicas incluem: a implementação seguiu o cronograma e o escopo definidos? Os recursos humanos, financeiros e materiais foram mobilizados como previsto? O público-alvo foi efetivamente atendido? Ao responder essas questões, esse tipo de avaliação permite identificar desvios entre o plano e a realidade, evidenciando possíveis falhas operacionais ou problemas de gestão.

Já a **avaliação de processos** tem como foco central investigar a qualidade da execução. Busca compreender como a intervenção está sendo implementada na prática, a partir de evidências sobre o funcionamento real do programa. Essa análise

[14]Este capítulo se baseia, de maneira geral, em diversas publicações relevantes da literatura especializada, incluindo Bamberger, Rugh e Mabry (2012), , Brasil (2018b), Rossi, Lipsey e Freeman (2004), Funnell e Rogers (2011), Gertler et al. (2016), Imas e Rist (2009), IJSN (2018d), Kusek e Rist (2004), Rogers e Woolcock (2023), UNFPA (2024), Comissão Europeia (2024), Patton (2008), OCDE (2023) e Vaessen et al. (2020), Neuman (2014).

abrange aspectos como a qualidade do atendimento, a satisfação dos beneficiários, a atuação das equipes de ponta e as adaptações feitas ao longo da execução. Também considera os mecanismos internos e o contexto que influencia a entrega: como as atividades foram realizadas? Que fatores facilitaram ou dificultaram a implementação? Como o programa é percebido por quem o executa e por quem o recebe? Ao responder essas questões, a avaliação de processos contribui para explicar os resultados observados — ou sua ausência —, conectando o modo como a política foi implementada aos efeitos alcançados.

A finalidade central desse tipo de avaliação é entender como a política está sendo executada na prática e identificar pontos de melhoria que possam fortalecer a intervenção. Ao analisar os processos operacionais, o alcance e a adesão do público-alvo, a avaliação de implementação fornece subsídios concretos para aprimorar a execução enquanto ela ainda está em curso. Realizar avaliações desse tipo é essencial para a gestão pública e para o aprimoramento das políticas. Afinal, uma boa implementação é condição necessária para que qualquer política gere os resultados esperados. Mesmo a melhor das ideias pode falhar se não for bem executada. Por isso, avaliações de implementação e de processos são ferramentas essenciais para apoiar a gestão pública na condução de políticas mais eficazes.

Do ponto de vista da gestão da política pública, esse tipo de avaliação traz diversos benefícios práticos:

- **Correções em tempo hábil**: Por serem geralmente conduzidas durante a execução, e não apenas ao final do ciclo, essas avaliações permitem identificar problemas operacionais ou desvios enquanto ainda há tempo de corrigi-los. Gestores podem, por exemplo, detectar que uma atividade prevista não está ocorrendo como esperado e intervir prontamente para ajustá-la, aumentando a probabilidade de o programa atingir suas metas.

- **Aprendizado organizacional**: A avaliação de processos promove uma cultura de aprendizado contínuo dentro da administração pública. Ao coletar feedback sistemático sobre a implementação, as equipes podem refletir sobre práticas, compartilhar lições aprendidas e aprimorar procedimentos, favorecendo o desenvolvimento institucional.

- **Melhoria da efetividade**: Ao identificar gargalos, atrasos, desperdícios ou fatores que estejam prejudicando a execução, abre-se espaço para ações corretivas que aumentem a efetividade do programa. Pequenos ajustes processuais, como readequações em cronogramas, melhorias na comunicação entre equipes ou reforço em capacitações, podem gerar ganhos expressivos de desempenho. Assim, a avaliação de processos funciona como uma ferramenta de gestão da qualidade, garantindo que os esforços sejam de fato convertidos em entregas reais.

2.1 Principais Perguntas Avaliativas

As avaliações de implementação e de processos buscam responder a perguntas-chave sobre a execução do programa. Em geral, essas perguntas investigam dimensões críticas da implementação, oferecendo uma visão abrangente sobre o funcionamento real da política.

Essas dimensões incluem, com frequência: **insumos, fidelidade, cobertura e qualidade, capacidade e fluxos operacionais, contexto, percepções de usuários e equipes, fatores de sucesso e desafios enfrentados**.

Conforme apresentado na Parte II deste guia, o ponto de partida para formular as perguntas avaliativas deve ser sempre a Teoria da Mudança (TdM) da política. Cada elo da cadeia lógica (insumos, atividades, produtos, resultados) pode ser aprofundado com base em perguntas que ajudem a entender se, e como, a implementação está ocorrendo. Além disso, é importante aplicar as **lentes transversais relevantes, como equidade, gênero, mudança transformacional, pegada ecológica, sustentabilidade**, entre outras, conforme o escopo e os objetivos da política.

Abaixo estão alguns exemplos típicos de dimensões frequentemente investigadas em avaliações de processos:

- **Insumos**: Os recursos necessários (financeiros, humanos, materiais) foram disponibilizados e gerenciados de forma adequada? Houve falta de recursos ou desperdícios que afetaram a implementação?
- **Fidelidade ao desenho, cobertura e qualidade**: O programa foi executado conforme o planejado? As atividades previstas ocorreram dentro do cronograma e na escala esperada? O programa atingiu seus produtos imediatos? A qualidade das entregas foi satisfatória? O público-alvo foi plenamente alcançado e recebeu os serviços previstos? Foram feitas adaptações ao longo da implementação? Em caso afirmativo, essas mudanças melhoraram a execução ou descaracterizaram o modelo original?
- **Capacidade e fluxos operacionais**: A equipe de implementação estava capacitada e em número suficiente? Os processos operacionais — como os fluxos de trabalho, mecanismos de coordenação, logística e sistemas de informação — funcionaram de forma eficaz? Foram identificados gargalos ou atrasos significativos?
- **Comparações contextuais**: Poucas implementações ocorrem em contextos homogêneos. Diferenças regionais, institucionais, culturais e temporais podem afetar significativamente a entrega do programa. Avaliações de processos devem identificar essas variações e analisá-las com cuidado. Houve diferenças relevantes na implementação entre regiões, municípios, ou grupos sociais? Quais lições podem ser extraídas dessas variações?

- **Percepções e qualidade**: Como usuários e equipes avaliam a qualidade dos serviços prestados? Há indicadores de satisfação, críticas recorrentes ou sugestões de melhoria?
- **Fatores de sucesso e desafios**: Quais fatores contribuíram para o bom andamento da implementação em determinados contextos? Que obstáculos foram enfrentados — como burocracia excessiva, resistência de atores locais, falhas logísticas ou instabilidade institucional — e como esses desafios afetaram a execução? Há risco de esses obstáculos comprometerem os resultados esperados?

A definição das dimensões a serem investigadas e das perguntas avaliativas deve ser personalizada a cada contexto, alinhada aos objetivos e à TdM da política em questão. Ao final, a avaliação de processos deve ser capaz de responder, com clareza: **"O que foi entregue, para quem, em que quantidade e com que qualidade?"** Essa resposta é fundamental, já que os resultados futuros de uma política pública dependem, em grande parte, da **qualidade da sua execução**.

2.2 Métodos e Abordagens: A Importância dos Métodos Mistos

Avaliações de implementação e de processos tipicamente se beneficiam de uma **abordagem metodológica mista**, que combina técnicas quantitativas e qualitativas de forma complementar. Essa estratégia é especialmente recomendada quando se busca mensurar o grau de execução — o "quanto" foi feito —, e entender os mecanismos, percepções e dinâmicas envolvidas na entrega da política — o "como" e o "por quê".

A literatura especializada destaca que os métodos quantitativos oferecem evidências objetivas sobre a realização de atividades, a cobertura de público e o cumprimento de metas. Já os métodos qualitativos são fundamentais para captar a experiência dos implementadores e beneficiários, compreender os fatores que influenciam a implementação e identificar variações contextuais relevantes. A integração dessas abordagens permite a triangulação de dados, reforça a validade dos achados e produz análises mais completas, que podem orientar melhor a tomada de decisões.

2.2.1 Fontes de Dados Quantitativos

As fontes de dados quantitativos são fundamentais para medir a extensão da implementação e monitorar a execução em tempo real. Em geral, essas fontes ajudam a quantificar o que foi realizado, para quem, com que frequência e em que volume.

- **Dados de monitoramento e registros administrativos**

Uma primeira fonte são os **dados de monitoramento e registros administrativos** do próprio programa. Esses dados incluem, por exemplo, o número de beneficiários atendidos, a frequência das atividades realizadas e a execução financeira em relação ao previsto. A análise de séries históricas e o acompanhamento de metas ajudam a estimar o grau de aderência da implementação ao planejamento. Em uma política de saúde, por exemplo, pode-se verificar quantas visitas domiciliares foram efetivamente realizadas em comparação às metas mensais ou anuais. Esses dados geralmente estão disponíveis em sistemas de informação gerencial, relatórios de execução ou bases de dados institucionais.

- **Questionários padronizados**

Outra fonte relevante são os **questionários padronizados**, aplicados tanto a beneficiários quanto a implementadores. Com eles, é possível coletar informações sobre a percepção dos usuários em relação à qualidade e acessibilidade dos serviços (por exemplo, satisfação com horário de funcionamento ou infraestrutura), assim como levantar a visão das equipes de execução sobre aspectos como capacitação recebida, sobrecarga de trabalho ou disponibilidade de materiais. As respostas podem ser organizadas em escalas quantitativas, como as escalas Likert, que captam o grau de concordância ou satisfação em uma escala ordinal, geralmente de 1 a 5.

- **Métricas de fidelidade e qualidade da implementação**

Além disso, muitas avaliações desenvolvem **métricas específicas de fidelidade ao desenho da intervenção e à qualidade da implementação**. Isso pode incluir, por exemplo, a proporção de componentes executados conforme o manual do programa ou o percentual de profissionais que participaram das capacitações previstas. Esses indicadores ajudam a verificar se os principais elementos do desenho foram efetivamente colocados em prática.

- **Observações estruturadas com checklists**

Por fim, as **observações estruturadas com checklists** são uma ferramenta importante para padronizar a coleta de dados durante visitas de campo. O avaliador pode, por exemplo, visitar escolas, unidades de saúde ou centros comunitários e registrar, por meio de uma lista de verificação, a presença de equipe técnica, a disponibilidade de materiais, a infraestrutura física e o cumprimento de rotinas previstas. Esses registros, quando padronizados, permitem construir indicadores quantitativos objetivos sobre diferentes dimensões da execução.

2.2.2 Fontes de Dados Qualitativos

As fontes qualitativas são fundamentais para compreender aspectos da implementação que os dados quantitativos não captam. Elas permitem interpretar contextos, identificar percepções e revelar dinâmicas institucionais ou sociais que influenciam a entrega do programa.

- **Entrevistas em profundidade**

Entrevistas semiestruturadas com atores-chave — como gestores, coordenadores, executores na ponta (professores, nutricionistas etc.) e beneficiários (alunos, pais) — possibilitam aprofundar dimensões da implementação que os números não explicam. A partir desses relatos, é possível colher percepções detalhadas sobre o funcionamento do programa, os desafios enfrentados e sugestões de melhoria. Por exemplo, se os indicadores apontam baixa execução em determinada região, uma entrevista com o coordenador local pode revelar que houve atrasos no repasse de recursos ou obstáculos políticos. Essas percepções qualitativas são essenciais para entender a complexidade contextual da implementação e explorar nuances como motivação da equipe ou articulação interinstitucional.

- **Grupos focais**

Reuniões com grupos de beneficiários ou implementadores promovem a escuta coletiva de experiências e percepções. Um grupo focal com beneficiários pode revelar, por exemplo, que a comunicação do programa não alcançou certas comunidades — um achado relevante para reorientar estratégias de divulgação. Já grupos com executores tendem a trazer à tona problemas operacionais cotidianos ou estratégias informais que ajudaram a viabilizar a execução no campo.

- **Observação direta (não estruturada)**

Além dos checklists padronizados, o avaliador pode realizar observações livres no ambiente de implementação. Observar a rotina de uma escola durante a entrega da merenda, por exemplo, pode revelar interações entre alunos e merendeiras, o clima do refeitório ou outros aspectos não previstos nos instrumentos formais. Esses registros ajudam a capturar dimensões mais subjetivas da execução.

- **Análise documental)**

A revisão de documentos é uma etapa complementar importante. Analisar manuais do programa, protocolos operacionais, contratos, relatórios narrativos, atas de reunião e legislações pertinentes permite entender o que foi planejado formalmente, contrastando com o que ocorreu na prática. Também ajuda a mapear decisões institucionais, mudanças de rumo e justificativas formais sobre ajustes realizados durante a execução.

2.2.3 Triangulação e Abordagem Integrada

O uso combinado de métodos quantitativos e qualitativos caracteriza a abordagem de **métodos mistos**, amplamente adotada em avaliações de implementação

e processos. A ideia central é realizar a triangulação das informações: verificar se evidências provenientes de diferentes fontes e métodos convergem para as mesmas conclusões ou se há divergências que merecem investigação adicional. Por exemplo, dados administrativos podem apontar alta execução de uma atividade, enquanto entrevistas com beneficiários revelam baixa satisfação com sua qualidade. Essa divergência pode indicar que a quantidade foi atingida, mas a qualidade ficou aquém do esperado. A integração dessas abordagens permite triangulação de dados, o que aumenta a validade das conclusões e contribui para análises mais robustas e orientadas à ação.

Outro aspecto metodológico importante nas avaliações de implementação é o caráter participativo do processo avaliativo. O envolvimento de implementadores na construção da avaliação — seja na definição das perguntas, na interpretação dos resultados ou na sugestão de soluções — aumenta a relevância dos achados e fortalece o compromisso institucional com a avaliação. Manter um diálogo contínuo com gestores ao longo do processo avaliativo também contribui para validar achados preliminares e fomentar o uso estratégico das evidências geradas (UNFPA, 2024; Comissão Europeia, 2024; Patton, 2008).

A literatura especializada reforça essa lógica de **complementaridade entre métodos**: dados quantitativos isolados podem apontar que algo não está funcionando, mas é ouvindo os envolvidos e analisando o contexto que se compreende por que isso ocorre. Inversamente, percepções qualitativas ganham força quando confirmadas por dados objetivos. Por isso, é comum que relatórios de avaliação de processos apresentem, lado a lado, análises quantitativas — como estatísticas descritivas sobre cobertura, execução financeira ou frequência de atividades — e achados qualitativos — como trechos de entrevistas, observações de campo ou percepções sobre a implementação.

2.3 Etapas para a avaliação de processos

Uma avaliação de processos bem conduzida exige planejamento rigoroso, fundamentação teórica sólida, escolhas metodológicas coerentes e uma abordagem colaborativa. Cada etapa — do desenho inicial à comunicação dos resultados — deve ser orientada para a produção de evidências úteis, contextualizadas e relevantes para a tomada de decisão. A seguir, apresentam-se princípios e etapas essenciais para estruturar avaliações de processos que efetivamente contribuam para o aprimoramento das políticas e programas analisados.

1. **Parta da Teoria da Mudança do programa**

   A avaliação deve ser desenhada com base na Teoria da Mudança (TdM), que explicita como a política ou programa pretende alcançar seus resultados. Uma TdM bem estruturada é o ponto de partida: ela guia a formulação das perguntas avaliativas e a seleção de indicadores, além de permitir compreender o que esperar da implementação em cada etapa.

2. **Defina perguntas avaliativas claras, factíveis e sensíveis às lentes transversais**

   As perguntas devem ser específicas, relevantes e viáveis, considerando os recursos disponíveis e o momento da avaliação. Perguntas bem formuladas ajudam a manter o foco da investigação e evitam a coleta de dados irrelevantes. Sempre que possível, envolva as partes interessadas na definição dessas perguntas — isso aumenta a utilidade e a legitimidade da avaliação para gestores, equipes técnicas e financiadores. Além disso, é importante aplicar lentes transversais relevantes ao formular as perguntas avaliativas, como equidade, gênero, sustentabilidade, pegada ecológica, mudança transformacional, entre outras, conforme o escopo e os objetivos da política. Essas perspectivas ajudam a capturar efeitos diferenciados e promover uma análise mais abrangente e inclusiva.

3. **Utilize métodos mistos e triangulação de fontes de dados**

   Adotar múltiplas fontes de evidência é essencial. Combine dados quantitativos (ex.: indicadores de execução) e qualitativos (ex.: entrevistas, observações) para captar diferentes dimensões da implementação. A triangulação — ou seja, a convergência de achados de diferentes métodos e fontes — aumenta a robustez das conclusões e permite uma análise mais contextualizada dos resultados observados.

4. **Envolva atores-chave no processo avaliativo**

   A participação de atores-chave — como gestores, técnicos de campo e beneficiários — enriquece a avaliação em todas as suas fases. Eles podem contribuir com o desenho da TdM, a formulação das perguntas, a interpretação dos dados e a validação dos achados. Esse envolvimento fortalece o compromisso institucional com a avaliação e aumenta as chances de uso efetivo das evidências. Por exemplo, discutir resultados preliminares com a equipe implementadora pode gerar soluções imediatas e fomentar uma cultura organizacional voltada para o aprendizado contínuo.

5. **Assegure rigor metodológico e qualidade dos dados**

   A credibilidade da avaliação depende da qualidade das informações utilizadas. Isso exige instrumentos bem desenhados, protocolos padronizados de coleta, amostras representativas (quando aplicável), capacitação das equipes de campo e respeito a princípios éticos — como confidencialidade e consentimento livre e informado dos participantes. Dados confiáveis são a base para conclusões úteis e decisões acertadas.

6. **Documente o processo de implementação e as mudanças ocorridas**

   Registre todas as adaptações, imprevistos e mudanças ao longo da implementação — como alterações de público-alvo, mudanças orçamentárias e políticas, ou reorganizações institucionais. Esses registros permitem contextualizar os achados e explicar desvios entre o planejado e o executado.

Também é importante documentar as decisões metodológicas tomadas durante a avaliação, garantindo transparência e reprodutibilidade.

7. **Comunique os resultados de forma clara e útil**

   Ao final da avaliação, os resultados devem ser apresentados de maneira acessível, com destaque para recomendações práticas e utilizáveis. Relatórios executivos concisos, oficinas devolutivas com gestores e visualizações claras dos dados são estratégias eficazes para garantir que as informações geradas orientem decisões e melhorias no programa.

2.4 Armadilhas a evitar

Avaliações de implementação e processos são fundamentais para entender como políticas e programas funcionam na prática, mas não estão livres de desafios metodológicos e analíticos. Alguns erros são recorrentes e podem comprometer a qualidade dos resultados, como a falta de triangulação de dados, a condução exclusivamente interna das análises ou a negligência quanto aos contextos locais. Reconhecer essas armadilhas — e saber como evitá-las — é essencial para produzir avaliações mais robustas, úteis e comprometidas com a melhoria contínua.

2.4.1 Avaliação de processos ou avaliação de resultados?

Alguns estudos misturam o escopo da avaliação de processos com o da avaliação de resultados. É preciso ter cuidado, pois cada tipo de avaliação tem uma finalidade própria e responde a perguntas distintas — por isso, o ideal é realizá-las separadamente: primeiro, uma avaliação de processos para entender se e como o programa foi implementado; depois, uma avaliação de resultados para analisar os efeitos gerados. Se forem conduzidas conjuntamente, é importante deixar claro o que se refere à implementação e o que se refere aos resultados, já que cada uma tem perguntas avaliativas distintas e pode exigir métodos diferentes. Misturar os focos pode comprometer a clareza e a utilidade da avaliação.

2.4.2 Fazer julgamentos precipitados sem considerar o contexto

Interpretar indicadores quantitativos sem compreender o contexto pode levar a conclusões equivocadas. Um resultado abaixo da meta pode ser causado por fatores externos, como greves, desastres naturais ou mudanças políticas. Por isso, é essencial combinar diferentes métodos e fontes para capturar a complexidade do cenário avaliado.

2.4.3 Sobrecarregar a coleta de dados

Planejar muitos instrumentos e indicadores na tentativa de cobrir todos os aspectos do programa pode sobrecarregar a equipe e os participantes, comprometendo a qualidade dos dados. O ideal é focar nas perguntas avaliativas essenciais e aproveitar os dados já disponíveis nos sistemas de monitoramento para evitar duplicações e desperdícios de recursos.

2.4.4 Ignorar vieses de avaliadores e informantes

Quando apenas membros da equipe interna conduzem a avaliação, há risco de viés de confirmação, ou seja, a tendência de buscar ou interpretar informações que reforcem percepções prévias sobre o sucesso do programa, deixando de lado evidências que apontem falhas ou desafios. Beneficiários, por sua vez, podem oferecer respostas socialmente desejáveis. Para mitigar esses riscos, é recomendável envolver avaliadores externos, garantir o anonimato nas pesquisas, triangular pesquisas com fontes objetivas e construir confiança junto aos participantes, deixando claro que a avaliação busca aprimorar, não punir.

2.4.5 Não planejar o uso dos resultados

Avaliações perdem relevância quando não há clareza sobre como os achados serão utilizados. A chamada "avaliação na gaveta" ocorre quando relatórios são produzidos, mas não influenciam decisões. Para evitar isso, é importante envolver os tomadores de decisão desde o início e prever desde cedo quando, como e com quem os resultados serão discutidos. Por exemplo, estabelecer que os achados da avaliação serão apresentados em uma reunião de planejamento anual pode aumentar a chance de sua incorporação efetiva.

Exemplo prático 20— Avaliando a implementação do Programa de Reforço da Alimentação Escolar

Após seis meses de execução do Programa de Reforço da Alimentação Escolar, a equipe de avaliação realizou uma análise de implementação e processos com base na sua Teoria da Mudança. O programa visava reforçar a alimentação nas escolas públicas do município, com o objetivo de aumentar a frequência escolar, os resultados nutricionais e educacionais dos alunos e reduzir a evasão.

Principais perguntas avaliativas:

- As refeições foram servidas conforme o planejado, em quantidade e qualidade adequadas?
- O público-alvo foi efetivamente alcançado? Houve variações territoriais?

- A logística de entrega, o preparo e a infraestrutura foram suficientes?
- Quais barreiras e soluções emergiram ao longo da implementação?
- Como os profissionais e beneficiários percebem a execução do programa?

Métodos e dados utilizados:

- Análise de dados administrativos sobre volume de refeições servidas e regularidade;
- Entrevistas com diretores, merendeiras, nutricionistas e coordenadores escolares;
- Grupos focais com professores e responsáveis;
- Observação direta com checklist estruturado em escolas;
- Questionários de percepção aplicados a pais e alunos.

Principais achados:

- 95% das escolas aderiram ao programa, com refeições servidas em 92% dos dias letivos, em média;
- Cobertura de cerca de 90% dos alunos presentes, com alta adesão;
- A maioria das escolas seguiu o cardápio nutricional previsto, com pequenas adaptações locais validadas por nutricionistas;
- As principais barreiras incluíram atrasos logísticos e infraestrutura precária em algumas unidades, contornadas com ajustes operacionais;
- Professores relataram melhora na rotina das salas de aula após a implementação do lanche reforçado, com alunos mais atentos e participativos;
- Famílias demonstraram alto grau de aceitação do programa, com relatos de alívio alimentar e valorização da escola;
- Nutricionistas recomendaram ampliar os canais de escuta entre escolas e coordenação técnica, para permitir ajustes finos nos cardápios conforme hábitos regionais.

- As ações educativas previstas com as famílias – como encontros informativos e envio de orientações para casa – foram implementadas de forma pontual e com baixa adesão, comprometendo a estratégia de integração entre escola e família no tema da alimentação saudável.

Recomendações práticas:

- Reforçar a logística de abastecimento e a infraestrutura de preparo em escolas com dificuldades;
- Ampliar o diálogo entre nutricionistas, merendeiras e gestores escolares para ajustes nos cardápios;
- Criar materiais de comunicação para orientar as famílias sobre alimentação complementar em casa e retomar os encontros educativos previstos no desenho do programa;
- Institucionalizar o monitoramento contínuo da implementação, com indicadores operacionais e escuta regular de usuários.

Fonte: FGV CLEAR, elaboração própria.

Aplicando na prática

Avaliações de Implementação e Processos

Boas práticas	**Armadilhas a evitar**
Parta da Teoria da Mudança do programa	Misturar avaliação de processos com avaliação de resultados sem clareza
Defina perguntas avaliativas claras, factíveis e com lentes transversais (equidade, gênero etc.)	Avaliar efeitos finais sem entender antes como o programa foi implementado
Use métodos mistos e triangulação para fortalecer os achados	Basear conclusões apenas em números, sem considerar o contexto
Envolva atores-chave desde o início da avaliação	Planejar coletas extensas e pouco focadas, que sobrecarregam equipes e respondentes
Assegure rigor metodológico e dados confiáveis	Ignorar vieses de avaliadores internos e respostas socialmente desejáveis
Documente adaptações e decisões metodológicas ao longo do processo	Produzir relatórios que não são usados por falta de planejamento prévio de uso
Comunique resultados de forma acessível e prática	Não envolver os tomadores de decisão na discussão dos achados

3 Introdução à Avaliação de Resultados[15]

A avaliação de resultados tem por objetivo verificar se uma política ou programa atingiu os efeitos pretendidos junto ao seu público-alvo, conforme definidos na Teoria da Mudança (TdM). Em outras palavras, busca responder: Até que ponto os objetivos iniciais foram alcançados na prática?.

Diferentemente da avaliação de processos, que analisa a forma como o programa é executado, a avaliação de resultados se concentra nas dimensões específicas que a intervenção pretende alcançar, verificando se as metas estabelecidas para cada dimensão foram ou não atingidas. Exemplos dessas dimensões incluem frequência escolar, acesso a serviços de saúde ou nível de renda das famílias.

É importante ressaltar que, embora a avaliação de resultados analise a eficácia de uma política em alcançar suas metas previstas, ela não busca provar causalidade– esse é o papel específico da avaliação de impacto. Dessa forma, seu foco é examinar se as mudanças pretendidas ocorreram nos indicadores de interesse após a intervenção. Resultados referem-se às mudanças observadas após a implementação da política, enquanto impactos dizem respeito às mudanças que podem ser atribuídas causalmente à intervenção, isolando o efeito da política de outros fatores externos.

Essa modalidade de avaliação pode ser realizada tanto durante a implementação da política pública, como estratégia contínua de aprendizado e ajuste, quanto ao final de um ciclo específico de execução, com foco em prestação de contas e decisões estratégicas sobre continuidade ou reformulação.

Como especificamos no plano de avaliação apresentado na Parte II deste guia, as perguntas avaliativas devem sempre partir da TdM, que oferece um referencial claro dos resultados esperados e das condições sob as quais se espera que eles ocorram.

Ao alinhar perguntas avaliativas com a TdM, o avaliador consegue examinar detalhadamente se cada resultado esperado foi alcançado e compreender melhor onde possíveis gargalos ocorreram. Por exemplo, se um produto previsto foi entregue, mas o resultado esperado não ocorreu, isso sugere a necessidade de revisar a lógica causal da intervenção. Desse modo, a TdM funciona como um mapa essencial que não apenas orienta a coleta e análise de dados, mas também possibilita validar ou ajustar as hipóteses iniciais do programa, aprimorando continuamente o desenho e a implementação.

[15] Este capítulo se baseia, de maneira geral, em diversas publicações relevantes da literatura especializada, incluindo Rossi, Lipsey e Freeman (2004), Patton (2008), Gertler et al. (2016) CDC (2017), Kusek e Rist (2004), Neuman (2014) e Imas e Rist (2009), Brasil (2018b), Vaessen et al. (2020), OCDE (2023), Sánchez (2024).

Quando a avaliação de resultados indica que determinado resultado não foi alcançado, duas direções analíticas se tornam possíveis: a existência de falhas na implementação ou a inadequação da lógica causal da política. Se uma avaliação de processo ou de implementação, conduzida em paralelo, não identificar problemas relevantes na execução — como desvios, baixa cobertura ou má qualidade das entregas —, o achado aponta para uma fragilidade na TdM. Nesse caso, é necessário revisar as hipóteses causais do programa, questionar a validade das conexões estabelecidas entre insumos, atividades, produtos e resultados, e, se necessário, reformular a estratégia da intervenção.

Conforme ressaltado na Parte II, é importante aplicar as **lentes transversais** – como equidade, gênero e inclusão, mudança transformacional, pegada ecológica, entre outras – no refinamento das perguntas avaliativas. Essa abordagem permite verificar não apenas se os resultados esperados foram alcançados em termos agregados, mas também em que medida os benefícios se distribuíram de forma justa entre diferentes grupos populacionais, se a intervenção promoveu mudanças estruturais ou sistêmicas relevantes, e se os resultados obtidos implicaram consequências ambientais adversas.

Perguntas avaliativas formuladas com essas lentes ajudam a identificar desigualdades, oportunidades para mudanças mais profundas e sustentáveis, além de efeitos ambientais relevantes da intervenção. Essas reflexões são essenciais para aprimorar o desenho e a implementação de políticas públicas, garantindo resultados mais equitativos, transformadores e ambientalmente responsáveis.

3.1 Finalidades para a Avaliação de Resultados

A avaliação de resultados pode cumprir diferentes finalidades ao longo do ciclo de políticas públicas, indo muito além da verificação do cumprimento de metas. Segundo a literatura especializada, destacam-se os seguintes usos:

- **Efetividade**: Verificar se as metas definidas foram efetivamente alcançadas, com base nos resultados esperados pela intervenção, conforme estabelecidos na TdM.
- **Prestação de contas**: Oferecer evidências concretas que justifiquem decisões estratégicas e reforcem a transparência diante de financiadores, gestores e sociedade.
- **Aprendizado e melhoria**: Identificar o que funcionou bem, o que não funcionou e por que, considerando, inclusive, os efeitos não previstos, a fim de orientar o aprimoramento de políticas em ciclos futuros.
- **Equidade e inclusão**: Avaliar se os resultados foram distribuídos de forma justa entre diferentes grupos sociais, com atenção especial a populações vulneráveis, gênero, raça e território.

- **Otimização de recursos**: Relacionar os resultados alcançados aos recursos utilizados, subsidiando decisões sobre eficiência e alocação futura de investimentos públicos.

Para que a avaliação de resultados alcance seus objetivos — como apoiar decisões, promover aprendizado e reforçar a prestação de contas —, é essencial combinar uma base metodológica consistente com estratégias que favoreçam o uso prático das evidências. Mais do que simplesmente apresentar os achados, a avaliação deve prever mecanismos concretos para que esses resultados sejam incorporados à gestão, tais como momentos de devolutiva, discussão com tomadores de decisão ou integração aos ciclos de planejamento.

3.2 Avaliação de resultados versus outros tipos de avaliação

Para compreender melhor o papel da avaliação de resultados, é importante diferenciá-la de outras modalidades de análise ex post, como a avaliação de impacto, a avaliação de processos e o monitoramento. A seguir, destacamos as principais distinções e relações entre essas abordagens.

3.2.1 Avaliação de Resultados e Avaliação de Impacto

A avaliação de resultados busca verificar se uma política ou programa atingiu as metas previstas e promoveu as mudanças esperadas nas dimensões estabelecidas na TdM. A pergunta central é: o programa alcançou os resultados que se propôs a atingir?

Já a avaliação de impacto vai além: procura identificar se as mudanças observadas podem ser atribuídas diretamente à intervenção, ou seja, se o programa foi a causa dos resultados encontrados.

Essa distinção tem implicações metodológicas importantes. A avaliação de resultados costuma utilizar métodos mais simples e acessíveis, como análises antes e depois, estudos de caso ou investigações qualitativas. A avaliação de impacto, por outro lado, exige técnicas mais robustas, como grupos de controle, experimentos ou quase-experimentos, que permitam inferir relações causais com maior segurança.

A seguir, apresentamos um quadro-resumo com as principais diferenças entre essas duas modalidades avaliativas:

Tabela 2: Quadro comparativo simplificado

Aspecto	Avaliação de Resultados	Avaliação de Impacto
Pergunta-chave	O programa alcançou os resultados previstos nas suas metas?	O programa causou as mudanças observadas?
Métodos típicos	Comparação antes-depois, comparação com metas, benchmarks, estudos de caso	RCTs, métodos quase-experimentais
Necessidade de grupo controle	Não	Sim

Fonte: FGV CLEAR, elaboração própria.

Exemplo prático 21— Avaliando o impacto e o resultado do Programa de Reforço da Alimentação Escolar

Vamos retomar o exemplo do Programa de Reforço da Alimentação Escolar, cujo objetivo é aumentar a frequência dos alunos. Uma avaliação de resultados poderia comparar a frequência escolar média antes e depois da implementação do programa, verificando se houve aumento neste indicador. Já uma avaliação de impacto buscaria identificar se esse aumento foi, de fato, causado pelo programa — por exemplo, comparando escolas participantes com um grupo de escolas semelhantes que não receberam a intervenção, controlando outros fatores que poderiam ter influenciado os resultados.

Fonte: FGV CLEAR, elaboração própria.

3.2.2 Avaliação de Resultados vs. Avaliação de Processos

A avaliação de resultados analisa o que ocorreu após a implementação de uma política, ou seja, se os resultados desejados foram de fato alcançados. Já a avaliação de processos foca no percurso: investiga como esses resultados foram construídos, observando a qualidade da execução e a entrega dos produtos e serviços previstos. As duas abordagens são complementares e, sempre que possível, devem ser realizadas de forma articulada. A avaliação de processos ajuda a interpretar os achados da avaliação de resultados, explicando por que certos efeitos ocorreram — ou por que não ocorreram.

Quando combinadas, oferecem uma visão mais ampla sobre o desempenho da política pública: a avaliação de resultados mostra o que foi alcançado; a de processos esclarece os caminhos, gargalos e fatores que influenciaram esses resultados.

Exemplo prático 22— Integração entre Avaliação de Resultados e Avaliação de Processos

Se um programa educacional não atinge as metas de aprendizagem previstas, uma avaliação de resultados poderá indicar que os níveis de desempenho dos alunos ficaram abaixo do esperado.

No entanto, uma avaliação de processos pode revelar as causas por trás desse resultado, como atrasos na entrega de materiais didáticos, baixa adesão dos professores às capacitações ou falhas na implementação de metodologias previstas.

Ao combinar essas duas abordagens, é possível compreender não apenas o que ocorreu, mas também por que ocorreu — fornecendo insumos mais robustos para o aprimoramento da política.

Fonte: FGV CLEAR, elaboração própria.

3.2.3 Avaliação de Resultados vs. Monitoramento

O monitoramento e a avaliação de resultados são práticas distintas, embora complementares. O monitoramento está centrado no acompanhamento contínuo da execução de uma política: coleta dados sobre o uso dos recursos, o cumprimento das atividades previstas e a entrega dos produtos, permitindo ajustes imediatos na gestão. Na medida do possível, o monitoramento também pode acompanhar os resultados de curto-prazo.

Já a avaliação de resultados se concentra nos efeitos concretos gerados pela política após sua implementação. O foco está em analisar se as mudanças esperadas — tanto as de curto prazo quanto as que se manifestam no médio e longo prazos, como frequência escolar, adesão a serviços ou renda das famílias — de fato ocorreram e em que magnitude.

Em geral, os dados de monitoramento são coletados em tempo real, durante a execução. Já os dados para avaliação de resultados são coletados após um período suficiente para que os efeitos esperados possam ser observados. Enquanto o monitoramento responde se a política está sendo implementada conforme o planejado, a avaliação de resultados analisa se essa implementação produziu os efeitos desejados.

Exemplo prático 23— Diferença entre Monitoramento e Avaliação de Resultados

No contexto do Programa de Reforço da Alimentação Escolar, o monitoramento pode acompanhar, por exemplo, o percentual de escolas que receberam a merenda dentro do prazo previsto — uma informação essencial para verificar se a implementação está ocorrendo conforme o planejado. Já a avaliação de resultados analisaria, em um mo-

mento posterior, se essa entrega regular de merenda contribuiu para aumentar a frequência escolar dos alunos ao longo de um semestre. Ou seja, enquanto o monitoramento observa o andamento da execução em tempo real, a avaliação de resultados examina se os efeitos esperados da política realmente se concretizaram.

Fonte: FGV CLEAR, elaboração própria.

3.3 Principais Perguntas Avaliativas

As perguntas avaliativas devem ser construídas com base na Teoria da Mudança (TdM) da política ou programa em análise. A TdM explicita quais resultados são esperados e em que sequência lógica eles devem ocorrer, servindo como referência para verificar se as transformações pretendidas foram de fato alcançadas.

Cada elo da cadeia lógica prevista na TdM oferece um ponto de partida para a formulação de perguntas específicas. A avaliação de resultados se concentra, principalmente, nos resultados imediatos e intermediários. Quando um resultado esperado não é observado, é essencial investigar possíveis explicações para esse desvio, como falhas na implementação, fatores contextuais ou hipóteses que não se confirmaram.

Resultados de longo prazo (impactos) também podem ser incluídos em uma avaliação de resultados, mas isso traz desafios adicionais. Sem uma avaliação de impacto, não é possível atribuir com confiança os efeitos observados à política ou programa avaliado. Impactos tendem a ser influenciados por diversos fatores externos e dinâmicas complexas. Por isso, a análise rigorosa dessas dimensões exige abordagens específicas voltadas à inferência causal, como as utilizadas em avaliações de impacto.

Além disso, é importante aplicar as lentes transversais que se fizerem relevantes no contexto da política avaliada — como equidade, gênero, mudança transformacional, pegada ecológica e justiça climática, conforme apresentado quando discutimos plano de monitoramento e avaliação. Essas lentes permitem aprofundar a análise e captar efeitos diferenciados entre grupos, estruturas ou territórios.

Exemplo prático 24— Aplicando lentes transversais: exemplo do Programa de Reforço da Alimentação Escolar

No contexto de um programa municipal voltado à oferta de merenda reforçada em escolas públicas, uma avaliação de resultados certamente mediria a mudança observada na frequência escolar. Adicionalmente, algumas perguntas avaliativas formuladas com lentes transversais poderiam incluir:

- Equidade e gênero: As melhorias na frequência escolar foram

homogêneas entre meninos e meninas? Houve variações entre regiões urbanas e rurais?

- Mudança transformacional: O programa contribuiu para transformar de forma duradoura a relação das famílias com a escola e a alimentação? Houve mudanças institucionais na forma como a gestão pública trata a segurança alimentar no ambiente escolar?
- Pegada ecológica: A expansão do fornecimento de refeições aumentou o volume de resíduos orgânicos ou embalagens? Foram adotadas estratégias de redução ou reaproveitamento?

Essas perguntas ajudaram a identificar não apenas se o programa alcançou sua meta principal (melhoria na frequência escolar), mas também quais foram seus efeitos indiretos ou complementares em termos sociais, institucionais e ambientais.

Fonte: FGV CLEAR, elaboração própria.

3.4 Métodos e Abordagens para Avaliação de Resultados

A avaliação de resultados pode recorrer a diferentes métodos e abordagens, a depender do tipo de resultado a ser analisado e da disponibilidade de dados. Embora não exija o mesmo grau de rigor metodológico que uma avaliação de impacto, é fundamental que utilize métodos sólidos, capazes de produzir evidências confiáveis sobre o alcance das metas previstas.

Nesse sentido, destaca-se a relevância da adoção de métodos mistos e da triangulação de fontes. A combinação de técnicas quantitativas e qualitativas permite não apenas verificar se os resultados foram atingidos, mas também compreender como e por que essas mudanças ocorreram. A triangulação contribui para aumentar a robustez e a credibilidade dos achados, ao validar as informações a partir de diferentes fontes e perspectivas.

3.4.1 Fontes e tipos de dados

Para uma avaliação de resultados consistente, é necessário mobilizar diferentes tipos de dados, dependendo da natureza dos indicadores e das perguntas avaliativas formuladas. As fontes mais frequentes incluem:

- **Dados quantitativos**

Indicadores numéricos, benchmarks e registros administrativos são centrais para mensurar os resultados de forma objetiva. Exemplos comuns incluem taxas de matrícula escolar, frequência em serviços de saúde, renda domiciliar e índices de insegurança alimentar.

- Indicadores numéricos permitem verificar com precisão o cumprimento das metas estabelecidas.
- Benchmarks — comparações com padrões, metas ou médias de referência — ajudam a contextualizar o desempenho observado.
- Registros administrativos costumam ser acessíveis, atualizados e de baixo custo para uso em avaliações.

- **Dados qualitativos**

São essenciais para capturar nuances, percepções e explicações sobre os resultados. São especialmente úteis para entender o contexto, os fatores que influenciam os resultados e os efeitos não previstos. Entre os métodos qualitativos mais utilizados, destacam-se:

- Entrevistas individuais, que aprofundam percepções de beneficiários, técnicos ou gestores;
- Grupos focais, que exploram visões compartilhadas por segmentos específicos da população;
- Estudos de caso, que permitem analisar em profundidade situações específicas e identificar fatores facilitadores ou barreiras à obtenção de resultados.

- **Dados primários e secundários: quando usar e como combinar**

  - **Dados secundários** são aqueles já existentes, coletados para outras finalidades — como registros administrativos, censos e pesquisas nacionais. São especialmente úteis para análises de maior escala e ajudam a economizar tempo e recursos.
  - **Dados primários** são coletados diretamente no âmbito da avaliação e permitem maior personalização. São indicados para captar informações qualitativas, contextuais e subjetivas, ou para mensurar dimensões que não estão disponíveis nas bases existentes.

O ideal é combinar dados primários e secundários sempre que possível. Fontes secundárias ajudam a verificar metas quantitativas, enquanto fontes primárias complementam com profundidade qualitativa e contexto. Essa combinação fornece uma base mais robusta, coerente e explicativa para os achados da avaliação.

3.4.2 Métodos mais utilizados na avaliação de resultados

A escolha dos métodos depende diretamente das perguntas avaliativas, dos resultados previstos na Teoria da Mudança e da disponibilidade de dados. Em avaliações de resultados, os métodos não buscam inferência causal rigorosa, mas devem produzir evidências consistentes sobre o alcance dos resultados esperados. Abaixo estão alguns dos métodos mais frequentemente utilizados:

- **Comparação antes e depois (pré e pós-intervenção)**

Consiste em comparar indicadores coletados antes e depois da implementação do programa, verificando a magnitude da mudança ocorrida no período. É um método direto e útil quando há uma linha de base confiável. Exemplo: comparar a frequência escolar média dos alunos antes da introdução de um programa de alimentação e seis meses após sua implementação.

- **Comparação com metas e benchmarks**

Os resultados alcançados são comparados com metas previamente definidas ou com médias de referência (nacional, regional, institucional). Isso permite avaliar se os resultados estão dentro ou acima do esperado, mesmo na ausência de linha de base. Exemplo: comparar o percentual de famílias que superaram a linha de pobreza após um programa de transferência de renda com a média nacional.

- **Análise de séries temporais e tendências**

Quando há dados periódicos ao longo do tempo, é possível observar tendências e verificar se houve mudanças no padrão dos indicadores após o início do programa. Essa abordagem é útil especialmente em políticas contínuas e programas de larga escala. Exemplo: observar a evolução da taxa de evasão escolar em uma região por cinco anos, antes e depois da ampliação da oferta de transporte escolar.

- **Estudos de caso**

Permitem analisar em profundidade contextos ou situações específicas. São especialmente úteis para identificar fatores que facilitam ou dificultam o alcance dos resultados em determinados territórios, populações ou instituições.

Exemplo: análise detalhada de uma escola que superou metas educacionais após a introdução de um programa de reforço nutricional.

3.4.3 Outras abordagens qualitativas úteis para contextos complexos

Além das abordagens apresentadas neste guia, há vários outros métodos que vêm ganhando destaque na avaliação de políticas públicas implementadas em contextos complexos, multissetoriais ou altamente dinâmicos, conforme discutido em Vaessen et al. (2020). Entre eles, destacam-se:

- **Process Tracing**: abordagem qualitativa que reconstrói, de forma sistemática, a cadeia de eventos entre uma intervenção e seus efeitos, buscando explicar como e por que determinados resultados foram (ou não) alcançados. Indicada em avaliações realizadas em contextos complexos e dinâmicos, em que múltiplos fatores interagem ao longo do tempo;
- **Qualitative Comparative Analysis (QCA)**: método comparativo que analisa múltiplos casos para identificar combinações de condições associadas ao sucesso ou fracasso de uma política. Em vez de buscar um único fator determinante, a QCA permite explorar diferentes caminhos que podem levar a um mesmo resultado, sendo especialmente útil em contextos com alta diversidade entre os casos analisados;
- **System Mapping e Dinâmica de Sistemas**: abordagens visuais e analíticas para mapear elementos de sistemas complexos, explorar interações entre políticas e atores, e modelar cenários e efeitos sistêmicos ao longo do tempo;
- **Social Network Analysis (SNA)**: permite mapear e analisar relações entre atores institucionais ou indivíduos, identificando padrões de colaboração, influência e fluxos de informação em redes de implementação ou governança;
- **Outcome Harvesting**: abordagem retrospectiva e participativa que parte da identificação de mudanças significativas ocorridas na realidade — positivas ou negativas — e investiga de que forma uma política ou intervenção contribuiu para essas mudanças, mesmo que elas não estivessem previstas inicialmente;
- **Most Significant Change (MSC)**: técnica narrativa e participativa de avaliação que coleta histórias de mudanças percebidas como mais significativas por diferentes grupos envolvidos na intervenção, promovendo aprendizado e diálogo qualitativo sobre efeitos transformadores.

3.4.4 Combinação de métodos: triangulação e coerência com a Teoria da Mudança

A triangulação de métodos e fontes permite verificar a consistência dos achados e ampliar a compreensão dos fenômenos avaliados. Sempre que possível, recomenda-

se combinar abordagens quantitativas (como análise de indicadores) com abordagens qualitativas (como entrevistas e estudos de caso), buscando convergência ou explicações para divergências entre fontes.

A escolha dos métodos deve ser coerente com os resultados esperados definidos na Teoria da Mudança, com o grau de profundidade desejado e com os recursos disponíveis. Em todos os casos, é essencial garantir a qualidade dos dados, a clareza das perguntas avaliativas e a utilidade prática dos achados para orientar decisões e ajustes na política.

Exemplo prático 25— Avaliação de Resultados: Programa de Reforço da Alimentação Escolar

Contexto:

Um município implementou um programa de reforço alimentar escolar com o objetivo de melhorar a frequência e reduzir a evasão, conforme previsto na Teoria da Mudança da política.

Perguntas avaliativas:

- Houve aumento da frequência escolar dos alunos após a implementação do programa?
- Em que magnitude e entre quais grupos o efeito foi mais expressivo?
- Os professores perceberam um maior engajamento da comunidade escolar com práticas alimentares saudáveis?

Dados utilizados:

- Dados quantitativos secundários: registros de frequência escolar mensal, antes e depois da implementação.
- Dados qualitativos primários: grupos focais com professores e entrevistas individuais com gestores escolares.

Resultados da avaliação:
A frequência escolar aumentou de 80% para 92% após seis meses do programa, superando a meta inicial de 90%. Professores participantes dos grupos focais relataram melhorias visíveis no engajamento dos alunos e famílias. As entrevistas com gestores apontaram que o fornecimento regular da alimentação contribuiu para reduzir atrasos e faltas recorrentes, especialmente entre alunos em situação de maior vulnerabilidade econômica e social.

A avaliação também identificou efeitos não previstos: algumas famílias relataram redução na alimentação em casa, o que levou à recomendação de reforçar as ações educativas junto às famílias, a fim de evitar uma substituição indesejada da alimentação doméstica pela escolar.

Recomendações práticas:

- Manter o monitoramento contínuo da frequência e atenção escolar como indicadores centrais do programa.
- Reforçar estratégias de comunicação com as famílias sobre a função complementar da merenda escolar.

Fonte: FGV CLEAR, elaboração própria.

Aplicando na prática

Avaliação de resultados

Boas práticas

- Defina desde o início o que será considerado sucesso, com base na Teoria da Mudança. Considere incluir lentes adicionais na análise.
- Utilize métodos mistos e triangule fontes de dados para fortalecer a credibilidade dos achados.
- Adapte a comunicação dos resultados aos diferentes públicos (gestores, técnicos, sociedade), usando formatos acessíveis.
- Registre não apenas os resultados esperados, mas também efeitos não previstos que geram aprendizados relevantes.

Armadilhas a evitar

- Avaliar apenas as médias de indicadores de resultados, sem considerar diferenças entre subgrupos (gênero, território, renda).
- Confundir avaliação de resultados com avaliação de impacto (causalidade) ou de processo (execução).
- Não planejar desde o início como os achados serão usados na gestão.

- Elaborar recomendações genéricas, sem vínculo direto com os achados ou com pouca viabilidade prática.

4 Avaliação de Impacto de Políticas Públicas[16]

A avaliação de impacto é uma ferramenta essencial para entender se uma política pública realmente causou as mudanças desejadas em determinada população. Ao contrário da avaliação de resultados, que verifica principalmente se os objetivos e as metas previstas foram atingidos após a implementação da política, e do monitoramento, que acompanha a execução das atividades e entrega dos produtos, a avaliação de impacto busca responder a uma pergunta mais desafiadora: as mudanças observadas ocorreram por causa da intervenção, ou teriam acontecido mesmo sem ela? Essa capacidade de estabelecer atribuição causal é fundamental para embasar decisões sobre expansão, reformulação ou descontinuidade de políticas públicas.

Em outras palavras, enquanto a avaliação de resultados mostra se os indicadores previstos melhoraram, a avaliação de impacto procura estabelecer se foi a política que provocou as melhorias nas dimensões que compõem os resultados e impactos da teoria da mudança. Trata-se, portanto, de uma análise de atribuição causal, que exige maior rigor técnico e critérios específicos de comparação.

Primeiramente, define-se impacto causal individual como a diferença entre dois resultados potenciais: O resultado em um indivíduo caso participe de uma política e o resultado para este mesmo indivíduo caso não participe da política. Para isso, é preciso contrastar os resultados observados com um cenário em que a política não tivesse sido implementada. Em outras palavras, é necessário estimar o que teria acontecido com os beneficiários na ausência da intervenção.

Esse cenário hipotético — chamado de contrafactual — nunca pode ser observado diretamente. Afinal, uma mesma pessoa não pode, ao mesmo tempo, ter participado e não ter participado de uma política pública. No entanto, pode-se inferir o impacto causal médio sobre os indivíduos beneficiários, ou sobre um subgrupo deles. Para isso, a avaliação de impacto busca estimar esse contrafactual da forma mais precisa possível, criando grupos comparáveis e controlando fatores externos que possam influenciar a participação no programa e os resultados de interesse.

Para alcançar esse objetivo, são utilizadas diferentes estratégias, como sorteio aleatório (experimentos), critérios objetivos de elegibilidade (regressão descontínua), comparações ao longo do tempo (diferença em diferenças), entre outros modelos estatísticos.

Esse tipo de avaliação é especialmente indicado quando se busca estimar os efeitos causais de uma política ou inovação — por exemplo, em contextos em que os custos de implementação são elevados e é fundamental verificar se os resultados

[16] As principais referências deste capítulo incluem Angrist e Pischke (2009), Gertler et al. (2016), Glewwe e Todd (2022), Imbens e Rubin (2015), White (2010).

justificam o investimento. Também é recomendável quando os efeitos não são imediatos, ou quando múltiplos fatores externos podem influenciar os resultados, dificultando a atribuição direta à intervenção. A avaliação de impacto é particularmente útil quando há incerteza se a política está, de fato, produzindo os resultados esperados, ou quando se busca gerar evidências robustas para orientar decisões estratégicas sobre expansão, reformulação ou descontinuidade da política avaliada.

Por exigir maior complexidade técnica e coleta de dados mais estruturada, a avaliação de impacto não substitui outras formas de avaliação, mas as complementa. Em um ecossistema de gestão orientada por evidências, cada modalidade tem um papel específico — e a avaliação de impacto ocupa o espaço da atribuição causal dos efeitos.

Este capítulo se apoia em referências consolidadas na literatura internacional e busca apresentar os conceitos fundamentais da avaliação de impacto com linguagem prática e acessível para quem atua no desenho, na gestão ou na avaliação de políticas públicas.

4.1 Entendendo a lógica da causalidade: o desafio do contrafactual

A essência da avaliação de impacto está em responder a uma pergunta simples, mas metodologicamente complexa: **quais mudanças realmente aconteceram por causa da política pública?** Nesse contexto, chamamos de tratamento a intervenção, programa ou política implementada que se deseja avaliar. Para responder com segurança, é necessário separar os efeitos da intervenção de tudo o que poderia ter ocorrido mesmo que a política não tivesse sido implementada.

Esse é o chamado **problema do contrafactual**: não é possível observar, ao mesmo tempo, o que teria acontecido com um mesmo grupo de pessoas com e sem a intervenção. Cada indivíduo, família ou território só vive uma dessas situações. O papel da avaliação de impacto é, portanto, **reconstruir o melhor cenário possível de comparação** — estimando, com métodos apropriados, o que teria ocorrido na ausência da política.

O impacto é definido como a diferença entre dois resultados:

- o resultado observado após a implementação da política, para quem foi exposto a ela (grupo de tratamento), e
- o resultado contrafactual — isto é, o que teria acontecido se a política não tivesse sido implementada.

Como esse segundo resultado nunca é diretamente observável para as mesmas pessoas, é preciso estimá-lo com base em **grupos de comparação válidos** — que sejam semelhantes aos beneficiários em todas as características relevantes, exceto por não terem recebido a política.

Essa construção é delicada. Se o grupo de comparação for sistematicamente diferente dos beneficiários — por exemplo, mais escolarizado, de outra região ou com renda mais alta — qualquer diferença nos resultados pode refletir essas características, e não o efeito da política. Esse é o chamado **viés de seleção**, um dos principais desafios da avaliação de impacto.

Para minimizar esse risco, diferentes estratégias são utilizadas, conforme o contexto e a disponibilidade de dados, com o objetivo de criar grupos comparáveis e fortalecer a credibilidade das estimativas. Essas estratégias metodológicas — como experimentos, quase-experimentos e outras técnicas estatísticas — permitem, sob algumas hipóteses, isolar o efeito da política e gerar evidências mais robustas sobre sua efetividade.

Essa reconstrução do contrafactual exige comparações válidas, baseadas em grupos semelhantes e métodos rigorosos. É nesse ponto que entram os conceitos de **validade interna** e **validade externa**, que ajudam a compreender tanto a força da inferência causal quanto a possibilidade de generalização dos resultados (ver Box 23).

Box 23 – Validade interna e validade externa: dois pilares da avaliação de impacto

Ao estimar o impacto de uma política, é fundamental considerar dois aspectos centrais da qualidade da evidência gerada:

- Validade interna diz respeito ao grau de confiança de que a diferença observada nos resultados foi realmente causada pela política — e não por outros fatores. Uma avaliação tem alta validade interna quando os grupos comparados são equivalentes em todas as características relevantes, exceto pela exposição à política. Em geral, métodos como experimentos aleatórios (RCTs) ou regressão descontínua (RDD) costumam apresentar alta validade interna, desde que bem desenhados e executados.

- Validade externa se refere ao quanto os resultados de uma avaliação podem ser generalizados para outros contextos, grupos ou momentos. Uma política pode ter mostrado efeito positivo em um piloto, por exemplo, mas isso não garante que os mesmos resultados ocorrerão ao ser expandida para outras regiões. A validade externa depende de fatores como a diversidade da amostra, o contexto da intervenção e o grau de representatividade do estudo.

Idealmente, uma avaliação de impacto busca combinar alta validade interna com boa validade externa — ou seja, gerar estimativas confiáveis que também sejam úteis para orientar decisões mais amplas, como a expansão da política.

Fonte: FGV CLEAR, elaboração própria.

4.1.1 Que efeito estamos medindo? Efeito Médio do Tratamento, Efeito Médio sobre os Tratados, Efeito Médio Local do Tratamento e outras estimativas possíveis

Ao realizar uma avaliação de impacto, é importante saber **exatamente qual efeito está sendo estimado**. Isso porque uma mesma política pode ter efeitos diferentes dependendo do grupo analisado, da forma como a intervenção foi implementada e do método utilizado.

Os quatro tipos de estimativas mais comuns são:

- **ATE – Efeito Médio do Tratamento (Average Treatment Effect)**

É o **efeito médio da política sobre toda a população elegível**. Responde à pergunta: se essa política fosse implementada para todos, qual seria o efeito médio esperado nos resultados?

Esse é o parâmetro mais amplo e, muitas vezes, o mais relevante para decisões de expansão da política. Porém, nem sempre é possível estimá-lo com segurança, especialmente em contextos em que apenas uma parte da população teve acesso à intervenção.

- **ATT – Efeito Médio sobre os Tratados (Average Treatment Effect on the Treated)**

É o **efeito da política sobre aqueles que, de fato, participaram dela**. Responde à pergunta: entre os beneficiários da política, qual foi o efeito médio observado?

Essa estimativa é especialmente útil quando o objetivo é avaliar se a política funcionou **para quem foi efetivamente atendido (beneficiários tratados)**. No entanto, ela não deve ser automaticamente extrapolada para a população geral, pois os participantes podem ter características específicas que influenciam os resultados.

- **LATE – Efeito Médio Local do Tratamento (Local Average Treatment Effect)**

É o efeito estimado para um **subgrupo específico de pessoas cuja participação foi influenciada por algum fator**, como um sorteio ou um corte de elegibilidade.

Por exemplo, em uma política que só atende famílias com renda mensal abaixo de determinado valor, o LATE representa o efeito médio **entre as famílias próximas desse limite**, que poderiam ou não ser beneficiárias dependendo de pequenas variações de renda. Esse é o parâmetro estimado por métodos como regressão descontínua ou variáveis instrumentais.

- **ITT – Efeito Médio de Intenção de Tratamento (Intention to Treat)**

É o efeito médio da política sobre o grupo de pessoas **designado a receber o tratamento**, independentemente de sua participação efetiva. Responde à pergunta: qual é o impacto médio da política entre aqueles que receberam a oferta de participar, independentemente de terem aderido ou não?

Em geral, é possível formar grupos de elegíveis e não elegíveis perfeitamente comparáveis — por exemplo, por meio de uma aleatorização entre os potenciais beneficiários. No entanto, nem sempre quem foi sorteado para participar do programa de fato participa, e algumas pessoas designadas para o grupo de controle acabam sendo atendidas. As ocorrências destes vazamentos ou não-conformidades são comuns na prática e podem comprometer a validade da comparação direta entre participantes e não participantes.

Nesses casos, o mais adequado é comparar os grupos designados ao tratamento com os designados ao controle, o que permite estimar com confiança o efeito causal da oferta do programa — ou da "intenção de tratar" (ITT) — mesmo diante de desvios de implementação.

Essa estimativa é especialmente relevante em análises de **custo-benefício**, pois reflete o impacto real da política tal como foi implementada.

4.2 Métodos Quantitativos de Avaliação de Impacto

4.2.1 Experimentos aleatórios (Randomized Controlled Trials – RCTs)

Os RCTs ou experimentos aleatórios são considerados o "padrão-ouro" da avaliação de impacto. A lógica é simples: a **alocação aleatória de quem recebe e quem não recebe a política cria grupos equivalentes** desde que a amostra seja suficientemente grande, eliminando o viés de seleção.

- Como funciona:

Indivíduos, escolas, comunidades ou outras unidades são sorteados aleatoriamente para receber (grupo de tratamento) ou não receber (grupo de controle[17]) a política. O sorteio é essencial porque distribui, de forma balanceada, todas as características — observáveis e não observáveis — entre os dois grupos. Isso faz com que eles sejam, em média, estatisticamente equivalentes no início do estudo, desde que a amostra seja suficientemente grande. Com essa equivalência assegurada, qualquer

[17] Embora os termos "grupo de controle" e "grupo de comparação" sejam frequentemente usados como sinônimos, há uma diferença conceitual importante: em avaliações experimentais, o grupo de controle é definido por sorteio e tende a ser estatisticamente equivalente ao grupo tratado; já em métodos não experimentais, na ausência de sorteio, utilizam-se grupos de comparação.

diferença observada nos resultados pode ser atribuída com confiança à intervenção analisada.

- Vantagens:
  - Alta validade interna: confiança de que os efeitos observados foram causados pela política.
  - Simplicidade na implementação e na interpretação dos resultados.
  - Possibilidade de estimar tanto ATT quanto ATE.

- Quando usar:
  - Quando há excesso de demanda pela política e não é possível atender a todos imediatamente — o sorteio se torna um critério justo de alocação.
  - Em intervenções-piloto para testar a efetividade antes de escalar.
  - Quando se deseja maximizar a credibilidade dos resultados, especialmente para decisões de expansão.

- Limitações:
  - Pode gerar questões éticas ou políticas, especialmente ao excluir parte do público potencial.
  - Exige planejamento prévio, linha de base e logística rigorosa.
  - Custo e tempo de implementação podem ser elevados.
  - Resultados podem ter validade limitada a contextos semelhantes.

4.2.2 Regressão Descontínua (Regression Discontinuity Design – RDD)

A regressão descontínua é um método robusto para estimar impacto quando a política utiliza **um critério objetivo de elegibilidade com ponto de corte**, como idade, renda, nota ou escore socioeconômico. A lógica é que indivíduos muito próximos desse corte são similares em tudo, exceto no fato de estarem acima ou abaixo do limite — o que permite usar essa descontinuidade como estratégia de identificação do impacto.

- Como funciona:

Compara-se o resultado médio de indivíduos logo acima do ponto de corte (elegíveis ou tratados) com o de indivíduos logo abaixo (não elegíveis). Se a única diferença relevante entre eles é a exposição à política, essa diferença pode ser atribuída ao efeito da intervenção.

Por exemplo, imagine um programa de bolsas estudantis concedidas apenas a alunos com nota acima de 7,0. Alunos que tiraram 6,9 e 7,1 são praticamente iguais em termos de características observáveis e não observáveis, mas apenas os que atingiram 7,0 recebem a bolsa. A regressão descontínua compara esses dois grupos quase idênticos e interpreta qualquer diferença nos resultados futuros como efeito da política — nesse caso, da bolsa.

- Vantagens:
  - Alta **validade interna**, próxima à de um experimento aleatório — especialmente quando a densidade de observações próximas ao corte é grande.
  - Usa critérios já adotados por programas reais (renda, idade, nota), o que facilita a aplicação.

- Quando usar:
  - Políticas que utilizam **critérios claros de elegibilidade numérica**, com um limite bem definido.
  - Quando há suspeita de que é possível obter **grupos muito parecidos dos dois lados do corte**.
  - Quando a política não pode ser sorteada, mas há regras de entrada objetivas.

- Limitações:
  - Estima um efeito local (LATE): válido apenas para unidades próximas ao ponto de corte.
  - Requer um número suficiente de observações ao redor do corte.
  - Exige testes para garantir que o critério de corte não foi manipulado.

4.2.3 Variáveis Instrumentais (Instrumental Variables – IV)

O método de variáveis instrumentais é utilizado quando há fatores não observáveis que afetam tanto a participação na política quanto os resultados, gerando viés de seleção que os métodos anteriores não conseguem corrigir. A estratégia consiste em encontrar uma variável externa (o instrumento) que afeta a participação na política, mas não afeta diretamente o resultado, exceto por meio da política.

- Como funciona:

A variável instrumental é usada para prever quem participa da política, em um primeiro estágio. Em seguida, estima-se o impacto apenas entre aqueles cuja participação foi influenciada por esse instrumento. O efeito estimado é chamado de efeito local médio do tratamento (LATE) — ou seja, o efeito médio para as pessoas cuja decisão de participar foi alterada pelo instrumento.

Por exemplo, imagine a avaliação de um programa de capacitação profissional. Pode ser que pessoas mais motivadas procurem espontaneamente o curso e, além disso, tenham melhores resultados no mercado de trabalho, independentemente do programa. Nesse caso, a motivação — que não é observada — gera viés. Se, por outro lado, o acesso ao curso for definido por sorteio (ou por distância geográfica até um centro de capacitação), essa variável pode funcionar como instrumento: afeta a participação, mas não afeta diretamente o desempenho no emprego, exceto por meio do curso.

- Vantagens:
  - Corrige viés de seleção por não observáveis, desde que o instrumento seja válido.
  - Pode ser aplicado em contextos em que não há sorteio nem pareamento possível.

- Quando usar:
  - Quando há viés de seleção não observável.
  - Quando existe alguma variação quase aleatória que afete a participação na política.
  - Quando o desenho da política permite usar fatores externos como fonte de variação plausível na participação.

- Limitações:
  - Encontrar um bom instrumento é difícil e requer forte justificativa teórica.
  - O efeito estimado é local (LATE) — válido apenas para o grupo de pessoas cuja participação foi alterada pelo instrumento.
  - A interpretação depende de pressupostos fortes, muitas vezes não testáveis.

4.2.4 Diferenças em Diferenças (Difference-in-Differences – DiD)

O método de Diferença em Diferenças (DiD) é uma das estratégias mais utilizadas em avaliações de impacto quando não há aleatorização. Ele parte do princípio de que, ao comparar as mudanças ao longo do tempo entre um grupo tratado e um grupo de comparação, é possível isolar o efeito da política.

- Como funciona:

Compara-se a variação nos resultados antes e depois da política no grupo tratado, subtraindo-se a variação no grupo de comparação no mesmo período. A ideia é que qualquer diferença adicional observada no grupo tratado reflete o impacto da intervenção, desde que os dois grupos tenham evoluído de forma paralela no passado[18].

Por exemplo, suponha que um estado adote uma política de tempo integral em algumas escolas a partir de 2022. Para estimar o impacto da medida sobre o desempenho dos alunos, comparamos a mudança nas notas entre 2021 e 2023 nas escolas que adotaram a política e em escolas semelhantes que continuaram com o modelo tradicional. Se ambas evoluíam de forma parecida antes da política, qualquer diferença adicional observada a partir de 2022 nas escolas de tempo integral pode ser interpretada como o efeito da intervenção.

- Vantagens:
  - Útil quando já existe um grupo de comparação sem intervenção, mas com trajetória similar no resultado de interesse antes da intervenção (pressuposto testável).
  - Pode ser aplicado com dados administrativos ou pesquisas recorrentes, o que facilita a avaliação de impacto com custos menores e em larga escala.
  - Capta tanto ATT quanto, em alguns casos, efeitos médios mais amplos, desde que o grupo de comparação seja representativo da população de interesse.

[18] Fórmula simplificada para cálculo do impacto a partir da estratégia de Diferenças em Diferenças (DiD):

Impacto = $(Y_{11} - Y_{10}) - (Y_{01} - Y_{00})$

Onde:

- Y_{11} = resultado do grupo tratado após a intervenção
- Y_{10} = resultado do grupo tratado antes da intervenção
- Y_{01} = resultado do grupo de comparação após a intervenção
- Y_{00} = resultado do grupo de comparação antes da intervenção

- Quando usar:
  - Quando a política foi implementada apenas em parte do território ou população.
  - Quando o tratamento é escalonado, ou seja, implementado em momentos diferentes para distintos grupos.
  - Quando há dados antes e depois da intervenção para grupos tratados e não tratados.
  - Quando não é possível realizar a alocação aleatória, mas existe um grupo controle plausível.

- Limitações:
  - Requer dados em pelo menos dois momentos no tempo (antes e depois da intervenção).
  - A validade depende da suposição de tendências paralelas entre os grupos — ou seja, que, na ausência da política, os dois grupos evoluiriam de forma semelhante.

4.2.5 Controle Sintético

O método de controle sintético é utilizado para estimar o impacto de políticas públicas quando há apenas uma (ou poucas) unidades tratadas –(como um município, estado ou país) e um conjunto de unidades não tratadas disponíveis para comparação. É indicado em contextos em que experimentos ou pareamentos diretos são inviáveis, mas existe um grupo de controle amplo o suficiente para construir uma comparação robusta.

- Como funciona:

Cria-se um ”grupo de controle sintético”ao combinar ponderadamente[19] várias unidades não tratadas, de forma que essa combinação reproduza o comportamento da unidade tratada antes da intervenção, tanto em características observáveis quanto na trajetória dos resultados. A ideia é que esse controle sintético represente de forma plausível o que teria acontecido com a unidade tratada na ausência da política. O impacto é estimado como a diferença entre os resultados observados na unidade tratada e os resultados estimados no controle sintético após a implementação da política.

[19] A combinação ponderada consiste em atribuir diferentes pesos às unidades do grupo de controle, de modo que a média ponderada de suas características e trajetórias anteriores se aproxime da unidade tratada.

Por exemplo, imagine que uma nova política de restrição de álcool no trânsito (como a Lei Seca) foi implementada no estado de São Paulo, mas não em outros estados. Para estimar o impacto dessa política sobre o número de acidentes de trânsito, é possível construir um "estado sintético" combinando dados de outros estados que não adotaram a política. Essa combinação ponderada forma um controle sintético que simula como São Paulo teria evoluído na ausência da intervenção. A diferença entre a trajetória real de São Paulo e a do estado sintético, após a implementação da política, pode ser interpretada como o impacto da Lei Seca.

- Vantagens:
  - Permite avaliação de impacto em casos com apenas uma (ou poucas) unidades tratadas.
  - Usa dados observacionais amplamente disponíveis.
  - Transparência na construção do contrafactual, com pesos atribuídos às unidades do grupo de controle.

- Quando usar:
  - Quando apenas uma unidade foi exposta à política e não há sorteio nem grupos de comparação equivalentes.
  - Quando se dispõe de dados longitudinais (séries temporais) de qualidade para a unidade tratada e para outras unidades comparáveis.
  - Em estudos de políticas implementadas em nível macro (como estadual ou nacional), especialmente reformas amplas, legislações ou mudanças institucionais.

- Limitações:
  - Requer série histórica longa e de boa qualidade antes da intervenção.
  - A seleção das variáveis preditoras e da janela de tempo pode influenciar fortemente os resultados.
  - Potencial dificuldade em encontrar unidades não tratadas suficientemente parecidas para compor um bom controle sintético.

4.2.6 Pareamento

O pareamento é uma estratégia que busca comparar beneficiários de uma política com indivíduos não beneficiários que compartilhem características observáveis semelhantes — como idade, renda, escolaridade ou localização. A ideia central é construir um grupo de comparação que seja, na medida do possível, equivalente

ao grupo tratado em todos os aspectos relevantes, exceto pela participação na política. Dessa forma, é possível estimar o impacto da intervenção ao comparar os resultados médios entre os dois grupos.

Entre as abordagens mais utilizadas, destaca-se o pareamento por escore de propensão (Propensity Score Matching – PSM), que calcula a probabilidade de participação na política com base em características observáveis e emparelha indivíduos com escores semelhantes. Outra estratégia é o pareamento direto em variáveis específicas, como idade ou renda. A escolha do método depende da qualidade dos dados e da complexidade do contexto avaliado.

- Como funciona:

Calcula-se, com base em um modelo estatístico, a **probabilidade de cada indivíduo ter participado da política** (escore de propensão). Depois, para cada beneficiário, identifica-se um ou mais não beneficiários com escores semelhantes, formando **pares ou grupos comparáveis**. Este método é o chamado pareamento por escore de propensão (ou propensity score matching).

Já no **pareamento direto**, o emparelhamento ocorre com base em variáveis específicas — como idade, renda ou escolaridade — sem a necessidade de gerar um escore único. Esse tipo de pareamento pode ser exato, quando os valores coincidem, ou aproximado, quando há uma margem de tolerância previamente definida. Essa abordagem tende a ser mais simples, mas exige atenção especial à escolha das variáveis e à disponibilidade de indivíduos com perfis semelhantes no grupo de comparação.

Por exemplo, suponha que um programa de qualificação profissional foi ofertado para jovens em situação de vulnerabilidade. Como não houve sorteio para definir quem participou, pode-se usar o pareamento para formar um grupo de comparação. No caso do pareamento por escore de propensão, calcula-se a probabilidade de participação na política com base em características como idade, escolaridade, renda e localização. Em seguida, cada jovem beneficiado é comparado a um não beneficiado com escore semelhante. Já no pareamento direto, o emparelhamento poderia ser feito usando uma ou duas variáveis-chave — por exemplo, selecionando jovens com a mesma idade e faixa de renda. Nos dois casos, a diferença média nos resultados (como inserção no mercado de trabalho) entre os grupos pareados pode ser interpretada como uma estimativa do efeito da política.

- Vantagens:
  - Usa dados observacionais amplamente disponíveis.
  - Intuitivo e relativamente fácil de aplicar com softwares estatísticos.
  - Útil para políticas que já foram implementadas e não puderam ser sorteadas.

- Quando usar:
  - Quando há um grupo de comparação plausível, mas a alocação não foi aleatória.
  - Quando se dispõe de dados ricos em variáveis explicativas, possibilitando reconstruir um contrafactual crível com base em observações existentes.

- Limitações:
  - O pareamento só garante validade se todas as variáveis relevantes para a decisão de participar forem observadas — o que raramente é o caso.
  - Não controla para fatores não observáveis (ex: motivação).
  - Requer grande volume de dados com qualidade e variabilidade suficientes.

4.3 Como interpretar os resultados?

Independentemente do método utilizado — seja um experimento aleatório, uma regressão descontínua ou uma análise de diferença em diferenças — todos eles têm algo em comum: no final, o que se obtém é uma estimativa numérica que representa o impacto da política avaliada. Em termos técnicos, esse número é o coeficiente estimado do modelo. Ele indica, em média, quanto a política mudou o resultado de interesse em relação ao que teria acontecido na ausência da intervenção.

Esse valor é o ponto de partida para a interpretação dos achados. Mas, para que ele seja útil na prática, é preciso ir além do número: é necessário entender o que ele representa, sua precisão, relevância e suas limitações.

A seguir, apresentamos os principais elementos que devem ser considerados ao se interpretar os resultados de uma avaliação de impacto.

4.3.1 Entendendo o valor estimado

O resultado principal de uma avaliação de impacto costuma ser um **efeito médio**. Por exemplo: "a política X aumentou a frequência escolar em 5 pontos percentuais" ou "o programa Y reduziu em 20% a incidência da doença Z".

Esse valor representa a **diferença estimada entre quem recebeu a política e quem não recebeu**, em média. É importante esclarecer se esse efeito se refere a toda a população elegível (ATE), apenas aos participantes (ATT), ou a um subgrupo específico (LATE).

Na maioria dos estudos práticos, o que se estima é o efeito sobre os tratados (ATT). Já em desenhos mais abrangentes, como experimentos aleatórios representativos, o efeito pode ser interpretado como aplicável à população-alvo.

Além disso, é importante interpretar o efeito estimado em relação à média do grupo de comparação. Um impacto de 5 pontos percentuais, por exemplo, pode ser modesto ou expressivo, dependendo se a média do grupo de controle era 90% ou 10%. Sem esse referencial, corre-se o risco de superestimar ou subestimar a importância prática do efeito observado.

4.3.2 Significância estatística e intervalos de confiança

Além do valor estimado, é essencial comunicar o grau de certeza estatística. Isso aparece normalmente por meio de um valor-p (p-value) ou de um intervalo de confiança (IC).

Um impacto estimado de +6 pontos percentuais com $p < 0,05$ significa que há menos de 5% de chance de esse efeito ter ocorrido por acaso, caso a política não tivesse gerado impacto algum. Já um intervalo de confiança de 95% entre +2 e +8 p.p. indica que, com 95% de confiança, o verdadeiro efeito da política está dentro desse intervalo.

É importante ficar atento: se o intervalo de confiança incluir o valor zero, isso significa que o efeito pode ser nulo — e, nesse caso, considera-se que o resultado não é estatisticamente significativo. Além disso, intervalos muito amplos indicam baixa precisão nas estimativas, o que pode ocorrer, por exemplo, quando o tamanho da amostra é pequeno.

4.3.3 Direção, magnitude e relevância prática

Nem todo resultado estatisticamente significativo[20] é relevante do ponto de vista da gestão pública. Um aumento de 10% na frequência escolar, por exemplo, pode ser estatisticamente significativo, mas talvez não justifique os custos envolvidos na política. Por outro lado, uma redução de 10% nas hospitalizações pode ter grande relevância prática, gerando economia e benefícios sociais importantes.

Interpretar resultados significa traduzir números para a prática: quantas pessoas foram beneficiadas? O quanto melhoraram? Isso justifica o investimento feito?

Comparar com metas ou benchmarks pode ajudar: se a meta era um aumento de 30% e a política gerou 10%, o resultado é modesto. Mas se o objetivo era reduzir um problema grave e o impacto foi grande, mesmo que abaixo da meta, isso pode ser uma evidência importante da relevância da intervenção.

[20] Um resultado é considerado estatisticamente significativo quando a probabilidade de ele ter ocorrido por acaso, caso a política não tivesse gerado nenhum efeito, é suficientemente baixa — geralmente abaixo de 10%. Isso significa que o intervalo de confiança do efeito estimado não inclui o valor zero.

4.3.4 Causalidade e linguagem apropriada

Se o desenho da avaliação for robusto, é possível usar linguagem causal direta: "a política causou", ou "o programa gerou tal efeito".

Essa segurança, porém, depende dos pressupostos do método. Sempre que houver dúvidas — por exemplo, risco de viés de seleção e comparabilidade fraca —, é importante ser transparente com o leitor e evitar afirmações fortes sobre causalidade.

Exemplo prático 26— Atenção aos pressupostos

Imagine que uma avaliação concluiu que um programa de alfabetização aumentou a nota média dos alunos em 20 pontos. O resultado é estatisticamente significativo, com intervalo de confiança entre 5 e 35 pontos. À primeira vista, parece um impacto positivo da política.

Essa conclusão, no entanto, só é válida se o principal pressuposto do método for verdadeiro: que nada mais relevante tenha mudado nas escolas ou regiões avaliadas durante o período. Suponha, por exemplo, que outra política educacional — como uma nova formação de professores — tenha sido implementada ao mesmo tempo nas mesmas escolas. Nesse caso, o efeito estimado pode estar captando os dois efeitos combinados, e não apenas o impacto do programa de alfabetização.

Ou seja, mesmo com um resultado estatisticamente robusto, a interpretação causal pode estar equivocada se mudanças externas relevantes não forem controladas ou monitoradas.

Fonte: FGV CLEAR, elaboração própria.

4.3.5 Efeitos heterogêneos

Um efeito médio pode esconder variações importantes entre grupos. Sempre que possível, é útil explorar efeitos heterogêneos por faixa etária, sexo, localidade, escolaridade, perfil socioeconômico e outros recortes relevantes. Essa abordagem permite incorporar diferentes lentes de análise à avaliação, como as perspectivas de gênero, inclusão social, equidade ou justiça climática.

Essas análises ajudam a entender para quem a política funcionou melhor ou pior, e podem orientar ajustes no desenho da política, redefinir o público prioritário ou aprimorar o foco das ações. No entanto, devem ser conduzidas com cautela estatística. Testar múltiplos subgrupos sem planejamento prévio pode gerar achados espúrios por acaso. O ideal é que essas comparações estejam previstas no plano de análise da avaliação, com hipóteses justificadas e interpretação cuidadosa dos resultados.

Exemplo prático 27— Interpretando resultados de uma avaliação de impacto

Interpretar resultados significa:

1. Apresentar a direção e magnitude do efeito, com unidade clara;
2. Indicar se o resultado é estatisticamente significativo (valor-p, IC);
3. Avaliar sua relevância prática, com base em contexto, metas e custos;
4. Comunicar com clareza e transparência qualquer limitação do método.

Exemplo de interpretação:

O Programa de Reforço da Alimentação Escolar aumentou a frequência média dos alunos em 6 pontos percentuais (IC95%: 2 a 10 p.p.), o que sugere um efeito positivo da intervenção sobre a assiduidade escolar. Essa estimativa é consistente com a hipótese de que a oferta regular de alimentação contribui para reduzir faltas, especialmente em contextos de vulnerabilidade. A interpretação assume que, no período analisado, não houve outras mudanças relevantes — como alterações no transporte escolar ou em calendários letivos — que tenham afetado os grupos de forma diferenciada. Essa suposição foi considerada plausível com base nas informações disponíveis.

Fonte: FGV CLEAR, elaboração própria.

4.4 Como escolher o método mais adequado

Escolher o método de avaliação de impacto é frequentemente visto como um desafio técnico. Na prática, porém, trata-se de uma decisão que exige bom senso, compreensão do contexto da política e uso estratégico das informações disponíveis.

Não é preciso dominar estatística avançada para escolher bem. Basta ter clareza sobre a pergunta que se quer responder, os dados que já existem (ou que podem ser coletados) e o quão confiável precisa ser a resposta. A seguir, apresentamos alguns aspectos que ajudam a orientar essa escolha.

4.4.1 Que tipo de pergunta você quer responder?

O primeiro passo é entender qual é a principal dúvida sobre a política. Saber o que você quer descobrir é o melhor ponto de partida para escolher o método.

Tabela 3: Correspondência entre o tipo de pergunta e o efeito estimado

Pergunta principal	Tipo de efeito estimado	Quando usar
A política funcionou para quem participou?	ATT – Efeito médio sobre os tratados	Quando o interesse está nos efeitos entre os beneficiários reais da política.
Qual seria o efeito se a política fosse aplicada a todos?	ATE – Efeito médio na população elegível	Quando se quer estimar o impacto geral de uma expansão da política.
A política teve efeito para quem estava na linha de corte da elegibilidade?	LATE – Efeito local	Quando há um critério objetivo de acesso (como escore, nota ou renda).

Fonte: FGV CLEAR, elaboração própria.

4.4.2 O quanto os grupos são diferentes entre si? (e o que dá para observar?)

A principal dificuldade em avaliar impacto é separar o que foi causado pela política das diferenças pré-existentes entre os grupos comparados. Por isso, sempre que viável, recomenda-se adotar experimentos aleatórios (RCTs), que garantem grupos semelhantes por meio do sorteio. Quando esse tipo de desenho não é possível, o ideal é escolher o método que ofereça maior credibilidade na comparação, avaliando cuidadosamente os pressupostos exigidos por cada abordagem.

Nesse processo, vale se perguntar: os beneficiários da política são realmente comparáveis aos não beneficiários? E, se não forem, essas diferenças são observáveis e podem ser controladas?

4.4.3 Que tipo de dados você tem (ou pode ter)?

A **Tabela 4** resume como diferentes tipos de dados disponíveis podem orientar a escolha do método de avaliação mais adequado, destacando também o que cada abordagem permite fazer.

Tabela 4: Do dado ao método: qual caminho seguir?

Tipo de dado disponível	O que isso permite fazer	Método potencial
Dados antes e depois da política	Comparar trajetórias temporais entre grupos	Diferença em Diferenças (DiD)
Informações detalhadas sobre as unidades (renda, idade, escolaridade, localização etc.)	Parear unidades semelhantes	Pareamento (PSM ou direto)
Ponto de corte objetivo para acesso à política (ex.: nota mínima, escore, renda)	Comparar quem está imediatamente acima e abaixo do limite	Regressão Discontínua (RDD)
Fonte de variação quase aleatória (ex.: distância, ordem de inscrição)	Isolar variação exógena na participação	Variáveis Instrumentais (IV)
Lista completa de elegíveis antes da implementação	Planejar sorteio justo entre os elegíveis com coleta de dados desde o início	Experimento Controlado (RCT)

Fonte: FGV CLEAR, elaboração própria.

4.4.4 Como escolher o método mais adequado na prática?

Um dos critérios mais importantes para escolher o método de avaliação de impacto é o estágio em que a política se encontra. Além de considerar a disponibilidade de dados e a possibilidade de formar grupos de comparação similares, o momento da política — se está em planejamento, piloto, expansão ou já amplamente implementada — define quais métodos são viáveis ou exequíveis. Avaliações experimentais, como o RCT, exigem planejamento prévio e controle sobre a implementação, enquanto métodos observacionais, como Diferença em Diferenças ou Pareamento, podem ser aplicados quando a política já está em andamento e há registros históricos confiáveis. A tabela a seguir resume essas possibilidades, discutindo os métodos mais viáveis com base na situação prática e no momento da política.

Tabela 5: Escolha do método com base no estágio da política

Situação prática	Momento da política	Métodos mais viáveis	Observações
A política ainda não foi implementada, mas está em fase de planejamento	Planejamento (pré-implementação)	RCT	Melhor momento para estruturar um experimento com sorteio prospectivo.
A política está começando em pequena escala e será expandida gradualmente	Piloto ou início da implementação	RCT, RDD	A expansão controlada permite planejar sorteios por fases (RCT) ou aplicar RDD quando há critério de elegibilidade objetivo (ex.: escore, nota).
O programa já está em andamento com bons dados administrativos	Implementação em curso	DiD, Pareamento	É possível usar dados existentes para construir grupo de comparação, desde que haja comparabilidade e dados anteriores à política.
Há limitação de tempo, equipe ou orçamento para avaliação	Qualquer fase	DiD, Pareamento	Métodos mais simples e menos custosos, úteis quando bem aplicados e com dados mínimos confiáveis.
Existe uma variação quase aleatória (ex: distância, ordem de inscrição, fila etc.)	Durante ou após implementação	Variáveis Instrumentais (IV)	Essa variação pode ser usada como instrumento válido para isolar o efeito causal, desde que os pressupostos sejam plausíveis.
Há apoio político e institucional para uma avaliação rigorosa	Piloto ou início da implementação	RCT, RDD	Contexto favorável para aplicar métodos com alta validade interna, se houver viabilidade operacional.
A política é sensível ou enfrenta resistência institucional ou política	Após implementação	DiD, IV, Pareamento	Métodos observacionais podem ser mais aceitáveis, pois não interferem na alocação e podem ser conduzidos com menor exposição pública.

Fonte: FGV CLEAR, elaboração própria.

4.5 Dados qualitativos e métodos mistos: por que são importantes na avaliação de impacto?

Embora a avaliação de impacto tenha como foco principal estimar efeitos causais com base em dados quantitativos, os dados qualitativos desempenham um papel fundamental desde as fases iniciais da avaliação. Técnicas como entrevistas, grupos focais, observações de campo e análise documental ajudam a compreender melhor o contexto, formular boas perguntas avaliativas e levantar hipóteses sobre os mecanismos que explicam os resultados.

Essas abordagens são especialmente úteis para:

- Conhecer o contexto local e adaptar os instrumentos de coleta de dados à realidade dos territórios;
- Delimitar as perguntas avaliativas e levantar mecanismos de mudança plausíveis;
- Captar como diferentes atores (beneficiários, gestores, técnicos) percebem a política e atribuem significados à sua implementação;
- Interpretar resultados inesperados ou heterogêneos, identificar barreiras operacionais e apoiar recomendações sensíveis à realidade local.

Essa compreensão do contexto envolve não apenas dados sobre infraestrutura ou arranjos institucionais, mas também as formas como a política é percebida e recebida por quem a vivencia. Essas dimensões, muitas vezes decisivas para o sucesso ou fracasso de uma intervenção, nem sempre aparecem nos dados numéricos.

Além disso, os dados qualitativos enriquecem a interpretação dos resultados da avaliação de impacto, sobretudo quando os efeitos observados são nulos, inconsistentes ou variados entre grupos. Ao articular a análise de impacto com investigações sobre implementação, processos e contexto, surgem insights valiosos — muitas vezes qualitativos — que fortalecem tanto o desenho quanto a interpretação dos achados.

Essa combinação de abordagens é conhecida como métodos mistos e permite construir avaliações mais completas, sensíveis às realidades locais e úteis para a gestão pública. Ela alia o rigor estatístico da inferência causal à profundidade da investigação qualitativa, ampliando a capacidade de entender não apenas se a política funciona, mas como, por que e para quem.

Na prática, ao adotar uma abordagem avaliativa ao longo do ciclo da política, é comum — e desejável — recorrer a métodos mistos. Dados qualitativos e avaliações de processo enriquecem a compreensão sobre a implementação e ajudam a interpretar os resultados da avaliação de impacto. Essa integração fortalece a utilidade das evidências geradas e apoia decisões mais informadas, contextualizadas e alinhadas à realidade vivida nos territórios.

4.6 Desenho e implementação da avaliação de impacto

Uma boa avaliação de impacto depende não apenas da escolha do método, mas também da forma como ela é desenhada e implementada. Planejar com antecedência, coletar os dados certos, monitorar a execução e lidar com imprevistos são passos essenciais para garantir a credibilidade dos resultados. A **Tabela 6** resume os principais elementos que influenciam a qualidade de uma avaliação e o que deve ser observado em cada etapa.

Tabela 6: Elementos-chave no desenho e implementação de avaliações de impacto

Etapa da avaliação	Pontos críticos a planejar	O que observar ou garantir na prática
Definição do desenho	Escolha do método, tipo de efeito (ATE, ATT, LATE), pergunta avaliativa, viabilidade do grupo de comparação.	O método escolhido é factível? A pergunta avaliativa está clara? Há grupos comparáveis viáveis?
Planejamento do levantamento/coleta de dados	Linha de base, calendário, questionário, indicadores, treinamento da equipe de campo, orçamento.	Existem dados ou serão coletados dados pré-implementação? A coleta (se houver) cobre as variáveis relevantes? Os instrumentos são sensíveis ao contexto, foram pré-testados?
Execução da política	Monitoramento da implementação, registro de desvios, qualidade da entrega, cobertura do público-alvo.	A política foi implementada como planejado? Houve vazamento para o grupo de comparação?
Acompanhamento da amostra	Estratégias para reduzir perdas (atrito), atualização de contatos, previsões para seguimento.	Houve perdas na amostra? Elas foram diferenciais entre grupos de tratamento e controle? A amostra final continua representativa?
Análise dos dados	Estratégia analítica, controle de viés, análise de heterogeneidade, testes de robustez e sensibilidade.	Os pressupostos do método foram testados? Houve análise de subgrupos e verificação da consistência dos resultados?
Interpretação dos resultados	Integração com teoria da mudança, análise de mecanismos, triangulação com dados qualitativos e de avaliação de processo.	Os resultados fazem sentido no contexto? A avaliação traz informações sobre como e por que a política funcionou (ou não)?

Fonte: FGV CLEAR, elaboração própria.

Exemplo prático 28— Avaliação de impacto: Programa de Reforço da Alimentação Escolar com RCT

Contexto

Um município implementou, em 2022, um programa de reforço na alimentação escolar com o objetivo de aumentar a frequência dos alunos, especialmente em escolas com alto índice de vulnerabilidade social.

A política foi inicialmente aplicada em metade das escolas da rede municipal. A seleção foi feita por sorteio entre todas as unidades escolares, o que permitiu adotar um desenho experimental (RCT) representativo da população-alvo. Antes da implementação, grupos focais com professores e gestores escolares ajudaram a adaptar os instrumentos de coleta e a formular hipóteses sobre os mecanismos esperados (por exemplo, melhora na atenção e no bem-estar dos alunos).

Perguntas avaliativas

- O fornecimento diário de refeições saudáveis aumentou a

frequência escolar dos alunos?

- Os efeitos foram maiores em escolas com maior vulnerabilidade social?
- Quais mecanismos explicam os resultados observados?

Dados e método

A avaliação combinou abordagens quantitativa e qualitativa:

- Dados administrativos mensais de frequência escolar ao longo de um ano letivo completo;
- Grupos focais e entrevistas com professores, realizados no início e no fim do ano, para explorar percepções sobre mudanças no comportamento dos alunos;
- Desenho experimental (RCT): as escolas foram sorteadas aleatoriamente entre os grupos tratado e controle, assegurando comparabilidade e permitindo estimar o efeito médio da política sobre a rede (ATE).

Resultados

A frequência escolar aumentou, em média, 6 pontos percentuais nas escolas com o programa, em comparação com aquelas ainda não atendidas. O intervalo de confiança de 95% foi de 2 a 10 p.p., indicando um efeito estatisticamente significativo.

Os dados qualitativos reforçaram a interpretação dos achados: professores relataram melhora na atenção dos alunos durante o turno da manhã, menor irritabilidade e maior participação nas atividades escolares — mecanismos plausíveis para o aumento da frequência. Os efeitos foram mais pronunciados em escolas com maior prevalência de insegurança alimentar, sugerindo um impacto heterogêneo relevante.

Recomendações práticas

- Considerar a expansão do programa para todas as escolas da rede, com prioridade para aquelas com maior vulnerabilidade social.
- Os relatos dos professores sugerem que fatores fora da escola, como hábitos alimentares no ambiente familiar, também afetam o comportamento dos alunos. Considerar ações complementares junto às famílias pode potencializar os efeitos da política.

Fonte: FGV CLEAR, elaboração própria.

Exemplo prático 29— Avaliação de impacto do programa de reforço da alimentação escolar - Diferença em Diferenças (DiD)

Contexto

Em outro município, o Programa de Reforço na Alimentação Escolar foi iniciado em 2022 em cerca de 60% das escolas municipais. As demais continuaram com o modelo padrão de alimentação, por razões operacionais e logísticas.

Embora a alocação não tenha sido feita por sorteio, o município já dispunha de dados administrativos padronizados de frequência escolar dos dois anos anteriores para todas as escolas. Esse contexto permitiu aplicar uma estratégia quase-experimental baseada em Diferença em Diferenças (DiD), comparando a variação nas escolas com e sem o programa, antes e depois da implementação.

Perguntas avaliativas

- O fornecimento diário de refeições saudáveis aumentou a frequência escolar dos alunos?
- Os efeitos foram maiores em escolas com maior vulnerabilidade social?
- Há indícios de que o efeito tenha sido influenciado por fatores contextuais ou padrões pré-existentes?

Dados e método

A avaliação combinou análise quantitativa com verificação de pressupostos:

- Dados administrativos mensais de frequência escolar entre 2021 e 2023, para todas as escolas da rede;
- Aplicação do método de Diferença em Diferenças (DiD), estimando a variação na frequência escolar ao longo do tempo entre escolas com e sem o programa;
- Análise exploratória dos dados de 2021 para verificar se a suposição de tendências paralelas parecia plausível no período anterior à política;
- Levantamento complementar com entrevistas com diretores de escolas e coordenadores de alimentação, para explorar possíveis fatores operacionais que afetaram a adesão ao programa.

A análise estimou o efeito médio da política sobre as escolas tratadas (ATT), assumindo que, na ausência da intervenção, sua trajetória de frequência teria seguido padrão semelhante ao das demais.

Resultados

A frequência escolar nas escolas com o programa aumentou, em média, 5,2 pontos percentuais a mais que nas escolas controle entre 2021 e 2023. O intervalo de confiança de 95% foi de 1,8 a 8,5 p.p., indicando um efeito estatisticamente significativo.

As entrevistas qualitativas reforçaram os achados quantitativos, apontando que as escolas com maior engajamento na oferta alimentar e atuação da equipe gestora apresentaram melhores resultados. Os efeitos foram mais pronunciados em escolas urbanas periféricas, onde a evasão era historicamente mais alta.

Recomendações práticas

- Expandir o programa para as escolas ainda não contempladas, priorizando áreas com maior risco de abandono escolar;
- Caso a expansão ocorra por meio de critérios objetivos (como escore de vulnerabilidade), considerar estudos futuros com métodos complementares, como Regressão Descontínua (RDD);
- Investir no engajamento das equipes escolares e na gestão da alimentação, aspectos apontados como fatores facilitadores pelos relatos qualitativos.

Fonte: FGV CLEAR, elaboração própria.

A seguir, a **Tabela 7** compara as principais diferenças entre os dois exemplos de avaliação de impacto do **Programa de Reforço da Alimentação Escolar**, um com **experimento aleatório (RCT)** e outro com **diferença em diferenças (DiD)**.

Tabela 7: Quadro comparativo de desenhos de avaliação para o mesmo programa

Elemento	Avaliação com RCT	Avaliação com DiD
Forma de implementação	Escolas sorteadas para tratamento e controle	Escolas tratadas por decisão operacional (não aleatória)
Momento da avaliação	Planejada antes da implementação (ex ante)	Iniciada após a implementação (ex post)
Tipo de comparação	Grupos equivalentes criados por sorteio	Comparação entre evolução dos grupos ao longo do tempo
Força da inferência causal	Alta (com forte validade interna)	Moderada (depende da validade da hipótese de tendências paralelas)
Dados necessários	Linha de base + acompanhamento pós-intervenção	Dados antes e depois da política para ambos os grupos
Estimativa gerada	ATE (Efeito Médio do Tratamento) na população elegível	ATT (Efeito Médio sobre os participantes)
Possíveis limitações	Custo, tempo, questões éticas ou políticas relacionadas ao sorteio	Viés se a hipótese de tendências paralelas não for plausível
Relevância para expansão	Resultados com alta validade interna, úteis para decisões de escala ou replicação	Resultados com validade moderada, úteis para informar ajustes ou expansão quando há dados históricos confiáveis
Exemplo de uso ideal	Pilotos, chamadas públicas com excesso de demanda	Políticas já em curso, com cobertura ampla e dados históricos confiáveis

Fonte: FGV CLEAR, elaboração própria.

Aplicando na prática

Boas práticas

- Definir claramente o que será considerado "sucesso" com base na Teoria da Mudança. Ter metas e indicadores bem definidos facilita a análise e a comunicação dos resultados.
- Escolher o método com base na pergunta, no contexto e nos dados disponíveis. Mais importante do que usar o método "perfeito" é ter um método adequado e bem aplicado.
- Triangular fontes e métodos sempre que possível. A combinação de dados quantitativos e qualitativos aumenta a credibilidade e a utilidade dos achados.
- Planejar a avaliação desde o início da política, para ampliar as possibilidades metodológicas e garantir uma linha de base sólida. Atenção ao tamanho da amostra e ao poder estatístico!

- Documentar claramente os pressupostos do método escolhido. Toda avaliação depende de premissas, como comparabilidade entre grupos ou tendências paralelas, que precisam ser explicitadas e verificadas.
- Aplique lentes transversais sempre que possível ao formular perguntas avaliativas, como equidade, gênero, raça, pegada ecológica e justiça climática. Essas lentes ajudam a entender quem se beneficia (ou não) da política, e se há impactos indesejados sobre o meio ambiente ou sobre grupos historicamente excluídos.
- Comunicar os resultados de forma clara e útil para os gestores e o público. Use linguagem acessível, destaque achados relevantes e apresente recomendações viáveis.
- Criar espaços para uso ativo dos achados. Boas avaliações não se encerram no relatório: vale pensar em diálogos com equipes técnicas, oficinas de uso, momentos de devolutiva e aprendizado institucional.

Armadilhas comuns

- Confundir avaliação de impacto com avaliação de resultados ou de processo. Cada tipo responde a uma pergunta diferente e exige abordagens distintas.
- Usar métodos sem verificar os pressupostos necessários. Por exemplo, aplicar DiD sem checar se as tendências eram paralelas antes da política.
- Analisar apenas o "antes e depois" do grupo tratado, sem grupo de comparação. Isso não permite estimar impacto de forma confiável.
- Não planejar o uso dos resultados. Avaliações só têm valor se forem usadas para aprender, ajustar ou decidir.
- Fazer recomendações genéricas e pouco úteis. Recomendações devem ser claras, realistas e ligadas aos achados da avaliação.
- Ignorar limitações do método na hora de comunicar os resultados. Usar linguagem causal em estudos com baixa validade interna pode induzir a interpretações equivocadas.

- Tratar os dados qualitativos como acessórios. Ignorar a riqueza das percepções e dos mecanismos identificados em campo pode limitar a compreensão dos resultados e das oportunidades de melhoria.

Box 24 – Como elaborar um bom Termo de Referência (TdR) para avaliação de impacto

Um TdR bem elaborado orienta com clareza o que será avaliado, com que finalidade, e quais são as expectativas metodológicas, operacionais e de uso dos resultados. Abaixo, pontos essenciais para estruturar o documento:

- Objetivo da avaliação

  Descreva brevemente a política ou programa a ser avaliado, o momento em que se encontra (piloto, expansão, ciclo completo) e o propósito da avaliação (ex.: prestação de contas, aprendizado, apoio à tomada de decisão).

- Perguntas avaliativas

  Indique os principais impactos que se pretende medir.

  Dica: use sua Teoria da Mudança como referência para formular perguntas alinhadas aos objetivos da política. Considere incluir análise de subgrupos e efeitos indiretos, se relevantes.

- Escopo da avaliação

  Delimite o período, território, populações e recortes que serão analisados. Indique se a avaliação deve considerar lentes transversais (ex.: equidade, gênero, pegada ecológica, justiça climática).

- Abordagem metodológica

  Apresente expectativas quanto ao tipo de método (RCT, DiD, RDD, IV), possíveis restrições operacionais ou éticas, e dados disponíveis (linha de base, registros administrativos, surveys).

  Dica: sinalize se será necessária coleta primária de dados, com desenvolvimento de questionários e a supervisão da coleta de dados. Inclua solicitação de métodos para controle de qualidade dos dados coletados. Especifique se há a expectativa de utilização de métodos mistos.

- Produtos esperados

Liste os principais produtos: relatórios técnicos e executivos, apresentações para diferentes públicos, visualizações de dados (tabelas, gráficos, dashboards), e recomendações práticas.

Dica: inclua formatos acessíveis para comunicação interna e externa.

- Cronograma e entregas

  Inclua um cronograma com marcos principais da avaliação, como:

  - Workshop inicial de validação da Teoria da Mudança, para definir coletivamente as dimensões de impacto;
  - Elaboração do plano de avaliação;
  - Instrumentos de coleta de dados;
  - Relatório de campo descrevendo a coleta de dados e o controle de qualidade dos dados;
  - Relatórios preliminares e final com a análise dos dados;
  - Entregas intermediárias e devolutivas com as equipes gestoras.
  - Dica: reserve tempo para etapas participativas e ajustes.

- Perfil da equipe

  Descreva as competências desejadas (ex: métodos quantitativos, políticas públicas, conhecimento setorial, diplomas de graduação/ pós-graduação) e, se necessário, experiência com o público ou território da política. Indique se haverá interação com equipes locais.

- Orçamento e contrapartidas

  Apresente o valor estimado e a forma de pagamento. Informe os apoios institucionais previstos (acesso a dados, apoio de campo, articulação local). Considere incluir cláusula sobre uso dos achados para fins de comunicação ou publicação científica. Isso pode afetar o custo da proposta.

Fonte: FGV CLEAR, elaboração própria.

5 Avaliação de Custo-Benefício e Custo-Efetividade[21]

Toda política pública envolve escolhas — e, quase sempre, essas escolhas significam priorizar alguns aspectos em detrimento de outros. Em um contexto de recursos limitados, tempo escasso e demandas urgentes, o gestor público precisa conseguir responder a duas perguntas: (i) os resultados de uma política justificam o investimento realizado? e (ii) dentre as políticas sugeridas, qual deve ser implementada?

É aí que entram as avaliações econômicas, especialmente a Avaliação de Custo-Efetividade (ACE) e a Avaliação de Custo-Benefício (ACB). Ambas ajudam a responder, de formas diferentes, uma mesma pergunta essencial: vale a pena?

Este capítulo apresenta os principais conceitos, aplicações e limitações dessas ferramentas, com base em referências centrais da literatura. O conteúdo mostra como a avaliação econômica pode apoiar decisões públicas mais racionais, transparentes e informadas.

5.1 Por que avaliar economicamente uma política?

Políticas públicas sempre têm custos — mas nem sempre produzem os resultados esperados. Mesmo quando funcionam, é possível que existam alternativas mais eficazes, mais baratas ou mais sustentáveis. Avaliar economicamente uma política é uma forma de:

- Estimar quanto custa gerar um determinado resultado;
- Comparar diferentes alternativas com base no que entregam e no que consomem;
- Identificar quais políticas têm maior retorno social;
- Justificar ou revisar alocações orçamentárias;
- Apoiar decisões de expansão, reformulação ou descontinuidade.

As avaliações econômicas também ajudam a comunicar resultados de forma clara para gestores, financiadores e a sociedade.

[21] Este capítulo se baseia em referências centrais da literatura internacional sobre avaliação econômica de políticas públicas, incluindo ADB (2013), Banco Mundial (2011), BEI (2013), Boardman et al. (2018), Comissão Europeia (2015), Dhaliwal et al. (2013), Wholey, Hatry e Newcomer (2010), WSIPP (2024), Menezes Filho e Pinto (2017).

5.2 ACE e ACB: qual a diferença?

A Avaliação de Custo-Efetividade (ACE) e a Avaliação de Custo-Benefício (ACB) são duas formas de análise econômica aplicadas a políticas públicas, que respondem a perguntas diferentes, partem de pressupostos distintos e são mais adequadas a contextos específicos. Ambas ajudam a analisar a relação entre custos e resultados de uma política, mas com abordagens distintas: a ACB permite concluir se os benefícios superam os custos, enquanto a ACE compara a eficiência relativa entre alternativas que visam o mesmo resultado.

Tabela 8: Comparando ACE e ACB

Avaliação	Pergunta central	Quando usar?
Custo-Efetividade (ACE)	Quanto custa alcançar uma unidade de resultado?	Quando os benefícios são importantes, mas difíceis de expressar em termos monetários
Custo-Benefício (ACB)	Os benefícios superam os custos, em termos monetários?	Quando é possível atribuir valor monetário aos principais resultados

Fonte: FGV CLEAR, elaboração própria.

5.2.1 O que a ACE responde?

A Análise de Custo-Efetividade (ACE) busca responder quanto custa alcançar um determinado resultado, sem a necessidade de converter esse resultado em valores monetários. Trata-se de uma abordagem apropriada quando os resultados da política são padronizados e comparáveis — como o número de alunos que completam o ensino médio, vidas salvas com uma política de saúde, ou toneladas de CO_2 evitadas.

Com base nesse raciocínio, a ACE responde perguntas como: qual política reduz mais a evasão escolar por real investido? Ou ainda, qual estratégia de atenção básica salva mais vidas a cada R$ 100 mil aplicados?

A ACE permite comparar a eficiência relativa entre alternativas que buscam o mesmo objetivo, mesmo que operem por caminhos diferentes.

5.2.2 O que a ACB responde?

A Análise de Custo-Benefício (ACB) parte de uma pergunta central: os benefícios gerados por uma política superam os seus custos? Para responder a essa questão, é necessário monetizar os principais efeitos da intervenção, o que permite realizar comparações entre alternativas com finalidades distintas.

A ACB é útil em contextos em que o orçamento é limitado e múltiplas políticas públicas competem por recursos, quando os resultados podem ser convertidos em valores econômicos — como aumento de renda, produtividade, ganhos fiscais ou redução de custos sociais —, e também quando a decisão envolve comparar políticas muito diferentes entre si.

Exemplo prático 30— ACE e ACB são complementares

Programa de Reforço na Alimentação Escolar

Um município deseja avaliar economicamente os resultados de um programa de alimentação escolar voltado a melhorar a frequência dos alunos.

- Pela ACE (Avaliação de Custo-Efetividade):

  Calcula-se quanto custa gerar uma unidade de resultado — neste caso, o número de alunos com frequência escolar regular.

  Exemplo: o programa gerou 500 novos alunos com frequência superior a 90%, ao custo adicional de R$ 400 mil.

  Resultado: R$ 800 por aluno adicional com presença regular.

- Pela ACB (Avaliação de Custo-Benefício):

  Estima-se o valor monetário dos benefícios gerados pelo aumento da frequência escolar — como ganhos futuros de escolaridade e produtividade — e compara-se com os custos do programa.

  Se os benefícios monetários forem maiores que os custos, a política é considerada viável do ponto de vista econômico.

Conclusão:

As duas abordagens se complementam: a ACE permite comparar alternativas com base na eficiência relativa, enquanto a ACB avalia se os benefícios totais justificam o investimento.

Fonte: FGV CLEAR, elaboração própria.

5.3 Quando usar cada abordagem?

ACE e ACB são abordagens úteis em diferentes momentos do processo decisório e respondem a perguntas distintas.

A ACE é indicada para comparar a eficiência entre alternativas que buscam o mesmo resultado, estimando quanto custa gerar uma unidade de efeito. Não responde se a política "vale a pena" economicamente, mas ajuda a escolher o caminho mais eficiente.

A ACB, por sua vez, permite avaliar se os benefícios de uma política superam seus custos, desde que seja possível monetizar os resultados e contar com estimativas confiáveis de impacto.

Na prática, costuma-se usar a ACE para comparar alternativas viáveis e a ACB para decidir sobre a implementação, expansão ou substituição de políticas. Quando bem aplicadas, as duas abordagens se complementam.

Tabela 9: Quando usar ACE e quando usar ACB?

Situação	Use ACE	Use ACB
Tipo de resultado	Resultados padronizados e comparáveis (ex.: alunos, vidas salvas)	Resultados que podem ser monetizados com confiança
Natureza da comparação	Entre alternativas com o mesmo objetivo	Entre políticas com objetivos diferentes
Facilidade de monetização dos benefícios	Difícil ou indesejável monetizar diretamente	Viável e desejável monetizar os principais efeitos
Objetivo da análise	Comparar eficiência relativa entre caminhos possíveis	Julgar se os benefícios superam os custos totais da política

Fonte: FGV CLEAR, elaboração própria.

5.4 Métricas principais da avaliação econômica

Uma avaliação econômica gera números, mas o mais importante é entender o que esses números significam e como podem apoiar decisões sobre priorização, escala ou reformulação de políticas públicas. Para isso, é fundamental que os benefícios e resultados considerados nessas métricas venham de uma boa avaliação de impacto, capaz de estimar com confiança o que foi efetivamente causado pela política.

A seguir, apresentamos as principais métricas utilizadas em análises de custo-efetividade (ACE) e custo-benefício (ACB), destacando como cada uma pode apoiar a comparação entre alternativas e a avaliação da viabilidade de diferentes investimentos.

Tabela 10: Quadro resumo de métricas em ACE e ACB

Métrica	Explicação
Razão Custo-Efetividade (RCE)	Custo total dividido pelo número de resultados alcançados. Mostra quanto custa gerar uma unidade de resultado.
Razão Benefício-Custo (RBC)	Benefícios monetários divididos pelos custos totais. Valores acima de 1 indicam que os benefícios superam os custos.
Valor Presente Líquido (VPL)	Diferença entre os benefícios e os custos, descontados ao valor presente. Representa o ganho monetário líquido da política. Quanto maior, melhor.

Fonte: FGV CLEAR, elaboração própria.

5.4.1 Custo por unidade de resultado (RCE)

É a métrica básica da ACE. Mostra quanto custou gerar um resultado adicional com a política — por exemplo, um aluno com presença regular, uma consulta realizada, uma vaga ocupada.

- **Como calcular (Razão Custo-Efetividade – RCE)**:
  - Dividir o custo total da política pelo número de unidades de resultado alcançadas;

Exemplo: Se o programa custou R$ 4 milhões e gerou 5.000 alunos adicionais com frequência escolar adequada, o custo por resultado é R$ 800 por aluno.

- **Como interpretar**:
  - RCE menor → maior eficiência (menor custo por unidade de resultado);
  - RCE maior → menor eficiência relativa;

Quando se utiliza a ACE para decidir entre duas políticas, deve-se priorizar aquela com menor razão de custo-efetividade (RCE). A ACE permite comparar alternativas que visam alcançar o mesmo tipo de resultado, mas não indica se uma política é vantajosa em termos absolutos. Além disso, é útil para estimar o montante de recursos necessário à expansão da política com vistas a atingir uma meta específica de resultados.

5.4.2 Razão Benefício-Custo (RBC)

A RBC compara os benefícios monetários gerados por uma política com seus custos totais. É uma das métricas centrais da Avaliação de Custo-Benefício (ACB) e permite avaliar se a política gera retorno econômico positivo.

- **Como calcular**:
  - Estimar os benefícios da política em valores monetários;
  - Dividir o valor total dos benefícios pelos custos totais da intervenção.

Exemplo: se os benefícios estimados são de R$ 2 milhões e os custos são de R$ 1 milhão, a RBC é de 2,0. Cada real investido gerou R$ 2 de retorno estimado.

- **Como interpretar**:
  - RBC maior que 1 → benefícios superam os custos.
  - RBC menor que 1 → os custos superam os benefícios.
  - RBC igual a 1 → política se paga exatamente.

Ao usar a ACB para comparar duas políticas, recomenda-se priorizar aquela com a maior razão benefício-custo (RBC), desde que essa razão seja maior do que 1. No entanto, é importante ter cautela quando as escalas das políticas são muito diferentes: uma intervenção de pequena escala pode ter RBC elevada, mas gerar impacto social total menor do que outra com RBC ligeiramente inferior, porém com maior cobertura. Além disso, a expansão de uma política nem sempre preserva a mesma eficiência observada em pequena escala, devido a custos marginais crescentes ou mudanças no perfil dos beneficiários.

5.4.3 Valor Presente Líquido (VPL)

O VPL mostra o retorno líquido de uma política pública, considerando que recursos disponíveis hoje têm mais valor do que no futuro. Ele representa a diferença entre os benefícios e os custos da política, trazidos ao valor presente por meio de uma taxa de desconto.

A **taxa de desconto** é um parâmetro utilizado para trazer valores futuros para o presente — ou seja, para calcular quanto vale hoje um benefício ou um custo que ocorrerá no futuro.

- **Como calcular**:
  - Projetar os fluxos de benefícios e custos ao longo do tempo;
  - Aplicar uma taxa de desconto (por exemplo, 5% ao ano);

  - Trazer todos os valores ao presente e subtrair os custos dos benefícios acumulados.

- **Como interpretar**:
  - VPL positivo → a política gera um ganho econômico líquido;
  - VPL negativo → os custos superam os benefícios;
  - VPL igual a zero → os benefícios cobrem exatamente os custos, sem ganho nem perda.

5.5 Como interpretar essas métricas?

As métricas de avaliação econômica ajudam a informar decisões, mas não devem ser interpretadas de forma isolada. Seu valor depende da qualidade dos dados utilizados, do contexto em que foram produzidas e das questões que se pretende responder.

É fundamental assegurar que os resultados utilizados nas análises provenham de uma avaliação de impacto robusta – isto é, que os efeitos estimados tenham sido realmente causados pela política. Além disso, é importante considerar se há efeitos heterogêneos ou desigualdades que ficam ocultas nas médias, mas que podem ter implicações relevantes para a equidade ou a justiça social.

A avaliação econômica deve ser encarada como uma ferramenta de apoio, para comparar alternativas e justificar escolhas, enquanto a decisão final deve levar em conta também fatores contextuais, de equidade e a viabilidade prática da política analisada.

5.6 Etapas de uma avaliação econômica

Fazer uma Avaliação de Custo-Efetividade (ACE) ou uma Avaliação de Custo-Benefício (ACB) pode parecer complexo à primeira vista, mas o processo pode ser organizado em etapas claras e replicáveis, desde que haja um mínimo de estrutura e dados disponíveis.

A seguir, apresentamos um passo a passo acessível para orientar o desenho e a execução de avaliações econômicas aplicadas a políticas públicas.

5.6.1 Definindo a política e a população avaliada

Antes de tudo, é preciso delimitar o objeto da avaliação: qual é a política (ou componente dela) que será analisado, quem é o público-alvo e qual é o escopo geográfico e temporal da análise. Use a Teoria da Mudança da política como guia para entender os resultados esperados e os caminhos causais entre atividades, produtos, resultados e impactos.

5.6.2 Mapeando os custos da política

Os custos econômicos de uma política são os custos dos usos de todos os recursos produtivos alocados para a sua elaboração e implementação. Para isso, é preciso levantar os custos relevantes, com base em registros administrativos, documentos contábeis, orçamentos e dados de mercado. Incluir:

- Custos diretos: salários, insumos, materiais, infraestrutura;
- Custos indiretos: suporte técnico, monitoramento, gestão;
- Custos de oportunidade: valor mais alto dos ganhos potenciais associados a outras ações que deixaram de ser realizadas com os mesmos recursos.
- Quando aplicável, custos fiscais indiretos, como o peso morto da arrecadação tributária.

A clareza sobre os diferentes tipos de custo é essencial para a credibilidade da avaliação. Em um programa de reforço escolar, por exemplo, os **custos diretos** podem incluir despesas com refeições, transporte e remuneração de monitores; os **custos indiretos** envolvem o tempo de planejamento dedicado pela equipe gestora. Há ainda o **custo de oportunidade**, representado pelo uso de salas que, de outra forma, poderiam ser destinadas a outras atividades educacionais de valor potencial. Por fim, deve-se considerar o **custo fiscal indireto**, como o impacto decorrente da necessidade de aumento de impostos para financiar o programa — o que pode gerar perdas adicionais de bem-estar, conhecidas como **perdas de peso morto**.

5.6.3 Estimando os efeitos da política — com base em avaliação de impacto

Uma avaliação econômica não serve para medir o impacto da política — ela usa o impacto como insumo. Por isso, antes de calcular o custo por resultado (como na ACE) ou monetizar benefícios (como na ACB), é essencial estimar os efeitos causais da intervenção.

Isso requer, sempre que possível, uma avaliação de impacto bem conduzida, com métodos robustos que permitam isolar os efeitos reais da política e evitar confundi-los com outros fatores externos.

Se a política ainda não tiver sido avaliada diretamente, há alternativas:

- Utilizar dados de políticas semelhantes com impacto já conhecido;
- Realizar uma avaliação de impacto própria, adaptada ao seu contexto;
- Utilizar parâmetros da literatura, como meta-análises, revisões sistemáticas ou bancos de evidência.

Neste último caso, o analista importa um valor de efeito estimado a partir de estudos anteriores (por exemplo, que programas semelhantes aumentaram a frequência escolar em 5 pontos percentuais) e aplica esse parâmetro à política em análise. Para isso, é fundamental justificar por que essa evidência externa é adequada ao seu caso, considerando fatores como público-alvo, tipo de intervenção, escala, contexto institucional e trajetória prévia.

Se não for possível garantir uma estimativa causal com razoável confiança, é importante ser transparente sobre as limitações e evitar conclusões apressadas. A avaliação econômica pode seguir em caráter exploratório, mas sempre deixando claro que os resultados devem ser interpretados com cautela.

5.6.4 Monetizando os efeitos

Essa etapa, exclusiva da Avaliação de Custo-Benefício (ACB), consiste em atribuir um valor monetário aos efeitos gerados por uma política pública — ou seja, transformar os benefícios observados em termos econômicos. Esse processo pode ser realizado por meio de dois tipos principais de valoração: objetiva e subjetiva.

A **valoração objetiva** utiliza preços de mercado para quantificar benefícios. Isso é possível quando os efeitos da política correspondem a resultados com valor econômico observável, como aumentos de renda futura, redução de custos com transporte, ampliação do tempo disponível para o trabalho, ou quando envolvem custos evitados, como hospitalizações que deixam de ocorrer, abandono escolar prevenido ou diminuição da criminalidade. Por exemplo, no caso de um programa de vacinação, pode-se considerar o valor pago por uma dose de vacina quando esta é adquirida diretamente no mercado.

Já a **valoração subjetiva** busca estimar o valor que os beneficiários atribuem a um bem ou serviço, mesmo que ele não tenha preço definido no mercado. Essa abordagem recorre a métodos específicos, como a disposição a pagar — que estima quanto uma pessoa estaria disposta a desembolsar por determinado benefício —, ou o valor estatístico da vida, usado em políticas de saúde, segurança ou meio ambiente para mensurar benefícios associados à redução de riscos. No mesmo exemplo da vacina, a valoração subjetiva corresponderia ao maior valor que o indivíduo estaria disposto a pagar por uma dose, ainda que não haja compra direta.

A monetização dos benefícios é metodologicamente exigente e exige atenção aos pressupostos adotados. Diferentes métodos podem levar a estimativas bastante distintas, o que reforça a importância de documentar com transparência as escolhas realizadas e justificar a adequação das fontes e abordagens ao contexto da política avaliada.

5.6.5 Cálculo da Métrica da Política

Nesta etapa, o caminho se define de acordo com o tipo de avaliação econômica adotada.

Na Avaliação de Custo-Efetividade (ACE), calcula-se a razão entre o custo total da política e o número de unidades de resultado alcançadas. O objetivo é estimar quanto custa, em média, gerar um resultado específico — como o custo por aluno adicional com frequência escolar regular.

Já na Avaliação de Custo-Benefício (ACB), após a monetização dos benefícios, o passo seguinte é ajustar todos os fluxos de custos e benefícios ao valor presente, por meio da aplicação de uma taxa de desconto social. Com isso, é possível calcular as principais métricas da ACB: o Valor Presente Líquido (VPL), que expressa o ganho econômico líquido gerado pela política, e a Razão Benefício-Custo (RBC), que mostra a proporção entre os benefícios obtidos e os custos totais do investimento.

5.6.6 Análise de sensibilidade

Nenhum número em uma avaliação econômica é totalmente "exato" — toda estimativa envolve algum grau de incerteza. Por isso, é fundamental testar a robustez dos resultados por meio da análise de sensibilidade.

Essa etapa consiste em simular variações nos parâmetros assumidos ou estimados com margem de erro, como a taxa de desconto, o valor monetário atribuído aos benefícios ou projeções de custo em contextos futuros. O objetivo é verificar se as conclusões da avaliação se mantêm quando esses valores são ajustados dentro de limites razoáveis.

Simulações com cenários alternativos — como variações otimistas, pessimistas ou intervalos de confiança — ajudam a entender o quão sensíveis são os resultados às suposições adotadas. Isso oferece uma base mais sólida para recomendações realistas, transparentes e bem fundamentadas.

5.6.7 Interprete os achados com foco na decisão

O principal objetivo da avaliação econômica é subsidiar decisões. Para cumprir esse papel, os resultados precisam ser interpretados com clareza, comunicados de forma acessível e articulados ao problema analisado, ao contexto institucional e às alternativas concretas disponíveis.

É fundamental indicar o grau de confiança nas estimativas e explicitar os pressupostos adotados ao longo da análise. As implicações práticas devem ser destacadas de forma direta, apontando os possíveis caminhos a partir dos achados.

Além da eficiência econômica, a interpretação dos resultados deve incorporar outros critérios relevantes para a decisão pública, como equidade, viabilidade política e prioridades sociais. Boas decisões raramente se baseiam apenas em cálculos de custo-benefício: elas envolvem também julgamentos normativos sobre o que é justo, desejável ou apropriado em determinado contexto.

Box 25 – Custo por produto é diferente de impacto: entenda a diferença

Em muitas situações, especialmente quando ainda não há dados sobre os efeitos de uma política, é possível — e útil — calcular o custo por unidade de produto entregue. Esse tipo de análise ajuda a entender a eficiência operacional da política, ou seja, quanto custa entregar determinado produto.

Exemplos comuns:

- Custo por aluno matriculado;
- Custo por cesta básica entregue;
- Custo por atendimento realizado;
- Custo por quilômetro de estrada pavimentado.

Esses indicadores são importantes para:

- Comparar diferentes modelos de entrega;
- Planejar a expansão do programa;
- Controlar custos e evitar desperdícios.

Mas é importante ter clareza:

→ Custo por produto não mostra se a política funcionou. Ele mostra quanto custa entregar algo, mas não se essa entrega gerou impacto na vida das pessoas. Para chegar a essa conclusão, é necessário estimar os resultados da política com métodos de avaliação de impacto.

Exemplo prático: Um programa de alimentação escolar custa R$ 200 por aluno atendido. Esse dado é útil para fins operacionais e de planejamento. Mas ele não diz se o programa aumentou a frequência ou reduziu a evasão. Isso só pode ser respondido com uma avaliação que compare alunos com e sem acesso ao programa.

Dica: sempre que divulgar custo por produto, deixe claro o que ele representa, evitando interpretá-lo como sinal de sucesso da política.

Fonte: FGV CLEAR, elaboração própria.

5.7 Limitações e desafios práticos

Avaliações econômicas, como a Análise de Custo-Benefício (ACB) e a Análise de Custo-Efetividade (ACE), são ferramentas valiosas para apoiar decisões em políti-

cas públicas. Elas ajudam a entender se os resultados alcançados justificam os recursos investidos, a comparar alternativas e a definir prioridades em contextos de orçamento limitado. No entanto, seus números não devem ser interpretados de forma isolada. Como qualquer instrumento analítico, essas avaliações carregam incertezas, pressupostos e escolhas metodológicas que influenciam diretamente os resultados.

Nem todos os benefícios de uma política podem ser monetizados com precisão. Elementos como dignidade, inclusão social, equidade ou confiança institucional dificilmente se traduzem em reais — mas isso não significa que sejam menos importantes. Quando não for possível atribuir um valor monetário a esses efeitos, é essencial reconhecê-los qualitativamente e considerá-los com transparência na análise. Além disso, políticas com bom desempenho médio podem esconder desigualdades entre grupos, beneficiando uns mais do que outros. Por isso, é importante analisar efeitos heterogêneos sempre que os dados permitirem, sobretudo quando a equidade é um dos objetivos da intervenção.

A taxa de desconto, usada para trazer custos e benefícios futuros ao presente, também é um fator sensível: taxas mais altas penalizam políticas com impacto de longo prazo, como as voltadas à infância, ao meio ambiente ou à prevenção. Não existe uma taxa única ou "correta", mas a escolha deve ser coerente com os objetivos da política pública e testada em diferentes cenários.

Outro ponto essencial é a qualidade dos dados. Nem sempre os custos estão claramente documentados, os impactos são bem estimados ou os valores monetários são atualizados. Quando isso acontece, a recomendação é ser transparente quanto às limitações, adotar suposições razoáveis e recorrer à análise de sensibilidade para verificar a robustez dos resultados. Avaliações econômicas podem — e devem — lidar com incertezas, desde que isso seja feito com clareza e responsabilidade.

É importante lembrar que uma política com bom custo-benefício não é automaticamente viável, assim como políticas com baixo retorno econômico podem ser justificadas por razões sociais, políticas ou históricas. Fatores como viabilidade institucional, compromisso político, aceitação social e justiça distributiva também pesam na tomada de decisão. Por isso, a avaliação econômica deve sempre ser combinada com outras abordagens, como escuta de usuários, análise de implementação, critérios de justiça social e visão estratégica de longo prazo.

Por fim, tão importante quanto calcular os indicadores é saber interpretar e comunicar os resultados com clareza. Tente comunicar de forma compreensível, adaptando a linguagem ao público e sempre contextualizando os achados com base nos objetivos da política analisada.

Avaliações econômicas são ferramentas de apoio à decisão, ou seja, não substituem o julgamento de gestores, nem as escolhas coletivas que definem o rumo de uma política pública. Usadas com rigor e bom senso, ajudam a tomar decisões mais transparentes, justificáveis e orientadas para resultados que realmente importam.

Aplicando na prática

Passo a passo para usar ACB ou ACE com qualidade

Se você está pensando em aplicar (ou contratar) uma Avaliação de Custo-Benefício (ACB) ou de Custo-Efetividade (ACE), siga este roteiro prático para garantir que ela seja útil, confiável e orientada para decisão:

1. Defina bem o que será avaliado

   → Quais são os resultados esperados, quem será afetado e qual o horizonte de tempo da política.

2. Tenha os custos totais organizados

   → Inclua custos diretos, indiretos e, se possível, custos de oportunidade.

3. Use efeitos estimados com rigor

   → Preferencialmente com base em avaliação de impacto robusta.

   → Quando não for possível, use parâmetros confiáveis da literatura — com ressalvas claras.

4. Escolha o tipo de análise mais adequado

   → ACB se os resultados forem monetizáveis.

   → ACE se forem mensuráveis, mas difíceis de monetizar.

5. Avalie variações entre grupos

   → Se os efeitos são diferentes por subgrupo (gênero, renda, território etc.), inclua essas diferenças na análise.

6. Aplique taxa de desconto se os efeitos forem ao longo do tempo

   → Teste diferentes cenários (curto, médio, longo prazo). Justifique a escolha da taxa de desconto.

7. Faça análise de sensibilidade

   → Varie os parâmetros principais e observe como isso afeta os resultados.

8. Apresente os achados de forma clara

   → Use linguagem acessível, destaque as implicações práticas e o que pode orientar decisões.

9. Comunique os resultados com clareza, incluindo limitações e suposições
10. Combine, sempre que pertinente, a avaliação econômica com critérios de equidade e justiça social

6 Avaliação Executiva[22]

A Avaliação Executiva é uma modalidade rápida e estratégica de avaliação ex post, concebida para apoiar gestores públicos na tomada de decisão com base em evidências, em tempo hábil. Diferentemente das avaliações aprofundadas, que exigem longos prazos e levantamentos extensivos, a Avaliação Executiva oferece uma visão panorâmica sobre o desempenho de uma política pública, concentrando-se em um conjunto estratégico de perguntas avaliativas e utilizando dados disponíveis ou de fácil obtenção.

Essa abordagem costuma ser adotada quando há urgência para a tomada de decisão — seja para redirecionar, ainda durante o ciclo de uma gestão, uma política que enfrenta dificuldades, avaliar sua efetividade antes de uma eventual expansão ou informar decisões orçamentárias em prazos curtos. Sua lógica é transversal: abrange o diagnóstico do problema, o desenho, a implementação e os efeitos da política, sem se aprofundar extensamente em cada uma dessas etapas.

Embora seja formalmente uma avaliação ex post — realizada após ou durante a implementação —, seu escopo é mais amplo do que o de avaliações tradicionais. Além de examinar a execução e os resultados alcançados, a Avaliação Executiva pode também investigar questões relacionadas à identificação do problema e ao desenho da política, sempre que isso for relevante para aquele contexto.

Uma característica distintiva dessa modalidade é o equilíbrio entre profundidade e abrangência. Por operar sob um horizonte temporal curto e buscar resultados analíticos rápidos, a Avaliação Executiva adota uma perspectiva ampla, abordando múltiplas dimensões de forma balanceada, mas sem aprofundar-se em todas com o mesmo grau de detalhe. Isso não implica descuido metodológico — ao contrário, mantém-se o rigor essencial, reconhecendo que, no tempo disponível, nem todas as questões poderão ser esgotadas. Informações já existentes e análises ágeis são utilizadas para gerar pistas robustas que orientem decisões imediatas.

Em certo sentido, a Avaliação Executiva complementa outras modalidades avaliativas: ao identificar sucessos ou problemas iniciais, pode indicar a necessidade de avaliações mais aprofundadas posteriormente, como avaliações de impacto e avaliações aprofundadas de processos, quando apropriado.

Outro aspecto importante é que a Avaliação Executiva tende a ser menos onerosa e mais flexível do que avaliações tradicionais. Ao privilegiar o uso de dados já existentes e recorrer, frequentemente, a equipes internas, seus custos são reduzidos e o processo de execução torna-se menos burocrático. Em contrapartida, essa flexibilidade e customização metodológica reduzem a padronização, o que significa que

[22]Baseado em abordagens consolidadas e guias desenvolvidos com apoio do FGV CLEAR. Referências incluem Banco Mundial (2012), Kusek & Rist (2004), Imas & Rist (2009), Rossi et al. (2004), Patton (2015), Neuman (2014), Williams (2022), além de documentos oficiais do DPME (2020), Tesouro Britânico (2020), ACE (s.d.), Cabo Ocidental (2021), Brasil (2018b) e IJSN (2018b), entre outros.

diferentes Avaliações Executivas podem não ser diretamente comparáveis entre si.

Como detalhado nos capítulos anteriores, cada modalidade de avaliação se concentra em uma etapa específica do ciclo da política pública. Já a Avaliação Executiva sobrevoa transversalmente essas diferentes dimensões, analisando brevemente cada etapa sempre que possível — sem, no entanto, substituir a necessidade de avaliações aprofundadas quando o contexto as exigir.

Este capítulo apoia-se em abordagens consolidadas na literatura e em experiências nacionais e internacionais de avaliação rápida, com foco na utilidade prática e no apoio direto à tomada de decisão.

Box 26 – Características centrais da Avaliação Executiva

- Agilidade: Condução rápida, com duração típica de no máximo 6 meses, adequada a janelas decisórias curtas.
- Escopo focado e flexível: Definição de um conjunto limitado de questões avaliativas estratégicas, adaptadas às necessidades específicas do contexto e da política.
- Uso intensivo de dados existentes: Predominância de fontes secundárias (monitoramento, registros administrativos, avaliações anteriores), com coletas primárias complementares apenas quando imprescindível.
- Síntese: Consolidação estruturada e sintética dos achados para compor diagnósticos ágeis e integrados.
- Engajamento de gestores e partes interessadas: Apesar do curto prazo para a realização da Avaliação Executiva, é útil envolver ativamente os gestores públicos, técnicos do programa e atores estratégicos desde a definição das perguntas avaliativas até a interpretação dos resultados, favorecendo o uso efetivo das evidências.

Fonte: FGV CLEAR, elaboração própria.

6.1 Situações em que a Avaliação Executiva é Recomendada

A Avaliação Executiva é recomendada em diferentes situações práticas que exigem diagnósticos ágeis e orientados à ação, tais como:

- **Janelas de decisão curtas**: Quando há necessidade de embasar decisões orçamentárias, ajustes estratégicos ou revisões de políticas públicas em prazos

de semanas ou poucos meses, incompatíveis com avaliações tradicionais.

- **Reposicionamento de programas em execução**: Quando surgem dúvidas sobre a efetividade, coerência ou viabilidade de políticas já em curso, permitindo ajustes táticos oportunos.
- **Crises e situações emergenciais**: Em situações como desastres naturais, pandemias ou crises econômicas, avaliações rápidas oferecem feedback imediato para orientar e corrigir respostas públicas.
- **Avaliações intermediárias**: Em políticas plurianuais ou contínuas, a Avaliação Executiva permite verificar o progresso e realizar ajustes antes da conclusão do ciclo.
- **Preparação para decisões de escala ou continuidade**: Antes de ampliar um projeto-piloto ou decidir pela continuidade de uma política, fornece evidências rápidas sobre a efetividade inicial e as condições de implementação.
- **Promoção de aprendizado institucional**: Utilizada para consolidar práticas de monitoramento e avaliação, estimular o uso de evidências e fortalecer a cultura de gestão orientada a resultados.
- **Pressão por prestação de contas**: Em cenários de alta visibilidade pública ou exigência de transparência, possibilita apresentar rapidamente evidências sobre a execução e os primeiros resultados.
- **Análise de viabilidade para avaliações mais robustas**: Funciona como triagem para avaliar se vale a pena investir em avaliações de impacto rigorosas, evitando dispêndios desnecessários.

6.2 Limites de Aplicação da Avaliação Executiva

Embora seu valor seja significativo, é importante reconhecer que a Avaliação Executiva se aplica a políticas já em execução — ou seja, trata-se de uma avaliação ex post, realizada durante ou após a implementação. Ela não substitui avaliações ex ante, voltadas a políticas ainda não iniciadas, nem avaliações ex post mais aprofundadas, como as avaliações de processos ou de impacto. Sua vocação está em aprimorar a política em tempo real, gerar evidências úteis com agilidade e apoiar decisões estratégicas, sem a pretensão de comprovação causal definitiva.

Ainda assim, há **limites** e **desafios importantes** que devem ser considerados:

- **Profundidade e rigor limitados**: A agilidade da avaliação impõe trade-offs metodológicos. Nem todas as perguntas poderão ser respondidas com o mesmo nível de detalhe. Aspectos complexos, como a mensuração de impacto causal, extrapolam a capacidade de uma avaliação de curta duração. Os métodos utilizados podem sacrificar alguma precisão ou abrangência, resultando em análises mais descritivas.

- **Comparabilidade e generalização limitadas**: Por serem altamente adaptadas ao contexto, as Avaliações Executivas tendem a utilizar metodologias e escopos específicos, o que dificulta a comparação entre diferentes avaliações ou a generalização de seus achados para outros programas ou territórios. Seus resultados são válidos para o tempo, o local e as perguntas avaliadas, mas podem perder força explicativa fora desse recorte. Essa limitação deve ser explicitada nos relatórios e considerada ao se utilizar os achados para formulações mais amplas de política.

- **Pressão de tempo e restrição à reflexão crítica**: A dinâmica acelerada da Avaliação Executiva pode reduzir o espaço para reflexão aprofundada por parte da equipe avaliadora. O cronograma enxuto muitas vezes exige decisões metodológicas rápidas e pode limitar iterações analíticas ou revisões cruzadas. Para mitigar esse risco, é recomendável prever momentos curtos, mas estratégicos, de pausa e revisão — como checkpoints internos ou leitura crítica por alguém externo ao processo — garantindo consistência lógica e maior robustez às conclusões.

- **Confiabilidade dos dados**: A forte dependência de informações secundárias torna a avaliação sensível à qualidade e disponibilidade dos dados existentes. Bases de monitoramento desatualizadas ou lacunas em indicadores-chave podem comprometer a robustez das conclusões. Há também o risco de viés de seleção, especialmente quando se recorre a consultas rápidas com atores-chave mais acessíveis.

- **Escopo potencialmente restrit**o: Por definição, a Avaliação Executiva foca em questões avaliativas estratégicas e predefinidas. Com isso, aspectos relevantes podem ser deixados de fora. A seleção do escopo exige discernimento técnico para evitar a omissão de fatores críticos, ainda que não seja possível aprofundar todas as dimensões.

- **Menor robustez para accountability formal**: Embora seja altamente útil para ajustes de gestão e aprendizado rápido, uma avaliação executiva isolada pode não oferecer evidência suficiente para decisões de maior impacto político ou jurídico — como a descontinuidade de um programa estruturante — sem validações complementares.

- **Desafios de coordenação e participação**: O cronograma acelerado pode gerar sobrecarga na equipe avaliadora e nos participantes. Entrevistas concentradas em pouco tempo, agendas apertadas e a exigência de rápida resposta institucional podem afetar a profundidade das informações e reduzir o engajamento.

- **Riscos de viés e baixa independência**: Ao envolver equipes internas e procedimentos acelerados, há maior risco de viés — como a tendência de confirmar percepções prévias ou consultar apenas atores próximos e favoráveis à política. A ausência de uma perspectiva externa crítica pode comprometer a credibilidade dos achados. Para mitigar esse risco, é recomendável triangular fontes e métodos, envolver diferentes tipos de atores-chave e, sempre

que possível, incorporar revisores externos ou sessões de validação com participantes independentes.

Em síntese, a Avaliação Executiva não substitui avaliações mais aprofundadas quando estas se mostram necessárias. No entanto, é uma ferramenta valiosa em contextos de restrição de tempo e recursos, pois oferece insumos imediatos para decisões oportunas. Idealmente, deve integrar uma estratégia avaliativa escalonada: seus achados podem apontar temas que exigem aprofundamento futuro — como a necessidade de atribuição de causalidade por meio de uma avaliação de impacto ou uma análise mais detalhada da qualidade dos serviços por meio de uma avaliação de processos. Por outro lado, os resultados podem também indicar que, naquele momento, não há necessidade de avaliações adicionais.

Quando bem conduzida, a Avaliação Executiva equilibra agilidade e rigor possível, gerando evidências úteis e orientadas à decisão que qualificam a tomada de decisão na gestão pública mesmo sob pressão de tempo.

Box 27 – Experiências práticas de Avaliações Executivas

Avaliações Executivas e outras modalidades de avaliações rápidas vêm sendo cada vez mais incorporadas por governos ao redor do mundo. Abaixo, reunimos exemplos nacionais e internacionais relevantes, que ilustram diferentes usos, formatos e finalidades dessas avaliações.

Brasil – Experiências Nacionais

- Governo Federal – CMAP

  O Conselho de Monitoramento e Avaliação de Políticas Públicas (CMAP) realizou dezenas de avaliações executivas anuais com duração média de 4 a 6 meses. Os relatórios apoiaram o aprimoramento de políticas em áreas como desenvolvimento social, educação e infraestrutura.

  https://www.gov.br/casacivil/pt-br/assuntos/monitoramento-e-avaliacao/cmap

- Minas Gerais, Espírito Santo e Ceará

  Estados brasileiros também vêm integrando avaliações rápidas aos seus sistemas de gestão.

  - Minas Gerais: avaliações rápidas inseridas na gestão por resultados.
    https://fjp.mg.gov.br/monitoramento-e-avaliacao/
  - Espírito Santo: institucionalização pelo Instituto Jones dos Santos Neves (IJSN).
    https://ijsn.es.gov.br

- Ceará: conduzidas pelo Centro de Análise de Dados e Avaliação de Políticas Públicas (CAPP/IPECE).
  https://www.ipece.ce.gov.br/

Experiências Internacionais

- México – CONEVAL

  Desde 2007, o CONEVAL realiza a Evaluación de Consistencia y Resultados para programas federais, com foco em desenho, gestão e resultados iniciais.

  https://www.coneval.org.mx

- Chile – Direção de Orçamentos (DIPRES, em espanhol)

  A Evaluación de Programas Gubernamentales existe desde 1997, com relatórios que alimentam o processo orçamentário anual.

  https://www.dipres.gob.cl

- Colômbia – Departamento Nacional de Planejamento (DNP)

  Implementou em 2008 a Evaluación Ejecutiva (E2+), com duração média de três meses.

  https://www.dnp.gov.co

- África do Sul – Departamento de Planejamento, Monitoramento e Avaliação (DPME, em inglês)

  Desde 2019, promove Rapid Evaluations com foco gerencial e duração entre 1 e 3 meses, especialmente voltadas a programas sociais e pilotos inovadores.

  https://www.dpme.gov.za

Fonte: FGV CLEAR, elaboração própria.

6.3 Etapas de uma Avaliação Executiva

A Avaliação Executiva segue um processo orientado para a entrega de evidências úteis em tempo hábil. Embora o formato exato varie conforme o contexto, a experiência acumulada indica um conjunto típico de etapas, cada uma com implicações metodológicas específicas. Esta seção apresenta essas etapas e discute, para cada uma, os principais fundamentos e opções metodológicas, com base nos conceitos

e referências apresentados ao longo deste guia.

6.3.1 Definição do escopo e das perguntas avaliativas

A primeira etapa de uma Avaliação Executiva consiste em delimitar claramente o que será avaliado, com qual finalidade, em que prazo e com que tipo de dado. Essa etapa é decisiva para que a avaliação tenha foco estratégico e seja exequível no tempo disponível.

As perguntas avaliativas devem derivar da Teoria da Mudança da política, conforme discutido na Parte II deste guia. É a partir dos elos causais da intervenção — por exemplo, a relação entre atividades e resultados, ou entre produtos e impacto esperado — que se estruturam as principais hipóteses a serem analisadas. Em avaliações executivas que buscam uma visão panorâmica, é comum que as perguntas se distribuam ao longo de diferentes pontos da cadeia lógica da política: o problema está bem formulado e ainda é pertinente? O desenho da intervenção faz sentido frente à situação que pretende enfrentar? A implementação está ocorrendo conforme o previsto? Há sinais de que os efeitos desejados começaram a emergir? Os recursos têm sido utilizados de forma eficiente?

Além disso, dependendo do contexto e dos objetivos da política, pode-se incluir lentes transversais relevantes, como equidade, gênero e sustentabilidade. Esses recortes devem estar articulados às perguntas avaliativas e contribuir para refinar o foco da análise, mesmo em avaliações com escopo enxuto.

No momento inicial da avaliação, a equipe também deve responder a um conjunto de perguntas operacionais que ajudam a definir o escopo com realismo. É fundamental esclarecer qual decisão a avaliação precisa apoiar, o que deve ser respondido com prioridade e quais são o tempo e os dados disponíveis para isso. Quanto mais claras e viáveis forem essas definições iniciais, maior a chance de sucesso da Avaliação Executiva — tanto em termos de foco estratégico quanto de viabilidade prática.

6.3.2 Coleta e análise rápida de evidências

A coleta e análise de dados na Avaliação Executiva devem ser realizadas em ciclos curtos, com foco no essencial. Aqui, prioriza-se o uso de dados secundários e de fácil acesso. Um dos pilares da Avaliação Executiva é o aproveitamento máximo de dados já existentes:

- Dados de monitoramento e indicadores gerenciais: Produzidos pelo próprio programa ou por outros órgãos com atribuições de monitoramento ou transparência, incluem informações como número de beneficiários atendidos, execução financeira e metas operacionais.
- Dados administrativos (ou de rotina): São registros gerados continuamente pela administração pública no curso da oferta de serviços ou da execução de

programas. Incluem tanto dados de gestão de programas específicos (como cadastros de beneficiários ou frequência escolar registrada no sistema da Secretaria de Educação) quanto bases administrativas setoriais mais amplas, como os dados do SUS ou do mercado de trabalho.

- Pesquisas e estudos anteriores: Linhas de base, diagnósticos, avaliações anteriores e literatura acadêmica podem ser mobilizados para contextualizar os achados e aumentar a qualidade interpretativa.

Ao focar nesses dados, a avaliação economiza tempo e recursos, mas é essencial verificar sua qualidade (atualidade, cobertura, integridade). Sempre que possível, recomenda-se combinar fontes distintas e cruzar evidências (ex.: registros administrativos com percepções qualitativas), promovendo triangulação metodológica.

Quando o uso exclusivo de dados existentes não é suficiente, pode-se complementar com coletas primárias ágeis, que devem ser selecionadas de acordo com as perguntas avaliativas e com o tempo disponível. Entre as opções mais comuns estão:

- Entrevistas semiestruturadas com informantes-chave: utilizadas para aprofundar questões específicas com gestores, especialistas ou atores estratégicos, oferecendo interpretações e informações contextuais relevantes que não estão registradas em bases de dados.
- Questionários rápidos (online ou presenciais): úteis para levantar informações estruturadas junto a públicos mais amplos. Podem aferir, por exemplo, a satisfação de usuários com a qualidade do serviço, a percepção de acesso ou o conhecimento sobre direitos e benefícios. Ferramentas digitais (formulários, aplicativos) facilitam a aplicação em pouco tempo.
- Grupos focais e oficinas participativas com implementadores e beneficiários: permitem levantar percepções e experiências de forma rápida e interativa, facilitando a identificação de barreiras, estratégias adaptativas e sugestões de melhoria. Podem ser realizados com equipes de campo, usuários do serviço ou gestores locais.
- Observação direta estruturada: realizada em campo, com uso de checklists, permite verificar aspectos práticos da implementação, como funcionamento de equipamentos, atendimento prestado, uso de materiais ou organização de espaços. É particularmente útil quando há dúvidas sobre a aderência da prática ao desenho previsto.

Essas técnicas devem ser planejadas com foco na agilidade: uso de roteiros prédefinidos, aplicação simultânea por equipes distintas e atenção concentrada nas perguntas avaliativas mais relevantes. Como regra, coleta-se apenas o indispensável para apoiar a tomada de decisão. Em geral, a colaboração da equipe gestora é essencial para viabilizar a coleta desses dados no tempo necessário.

Para a análise de dados, recomenda-se o uso de métodos mistos (quantitativos e qualitativos) e sua triangulação. A abordagem de métodos mistos permite combinar a objetividade de indicadores numéricos com a riqueza interpretativa de dados qualitativos, fortalecendo a validade dos achados e ampliando sua utilidade para a gestão.

Por exemplo, dados quantitativos extraídos de sistemas administrativos podem mostrar uma alta taxa de cobertura de uma política, enquanto entrevistas com implementadores ou beneficiários revelam dificuldades operacionais, baixa adesão real ou insatisfações com a entrega do serviço. Quando convergentes, essas evidências reforçam as conclusões da avaliação; quando divergentes, ajudam a identificar contradições relevantes que merecem investigação mais profunda.

A triangulação de métodos é, portanto, uma estratégia eficaz para fortalecer o uso de evidências em avaliações rápidas, permitindo compor diagnósticos mais equilibrados e fundamentados, mesmo sob restrições de tempo. Os dados quantitativos oferecem medidas objetivas sobre a execução e os resultados (como taxas de cobertura e execução financeira), enquanto os qualitativos ajudam a entender mecanismos e percepções — como experiências de usuários e explicações de gestores sobre dificuldades.

Essa combinação aumenta a robustez dos achados e qualifica a interpretação, contribuindo para decisões mais bem informadas mesmo em contextos de tempo limitado.

Tabela 11: Fontes e tipos de dados

Tipo de dado	Exemplos
Secundários (priorizados)	Dados administrativos, bases de monitoramento, relatórios internos
Primários (complementares)	Entrevistas rápidas, oficinas, questionários, observação direta
Fontes documentais	Leis, portarias, contratos, atas, planos, materiais de comunicação

Fonte: FGV CLEAR, elaboração própria com base em Neuman (2014), Imas e Rist (2009) e Kusek e Rist (2004).

6.3.3 Síntese dos resultados e comunicação para a utilidade

A etapa final da Avaliação Executiva consiste na sistematização crítica dos achados e na sua comunicação de forma clara, objetiva e voltada ao uso prático. É nesta fase que os dados brutos e as informações dispersas se transformam em diagnóstico estruturado e relevante para a tomada de decisão.

A síntese deve interpretar os dados à luz das perguntas avaliativas, articulando evidências quantitativas e qualitativas em uma narrativa coerente sobre o desempenho da política. Esse processo inclui a identificação dos principais achados, a

formulação de conclusões objetivas e, quando pertinente, a proposição de recomendações acionáveis. O foco está em destacar implicações práticas e lições aprendidas, com linguagem acessível e orientação ao uso dos achados da avaliação.

A comunicação dos resultados deve considerar o perfil e as necessidades do público-alvo. Em geral, combina-se diferentes formatos de entrega, como:

- Relatórios sintéticos e orientados à decisão;
- Apresentações presenciais para gestores, com foco nos principais achados e recomendações;
- Notas técnicas e sumários executivos;
- Dashboards, infográficos ou slides para apresentação rápida;
- Oficinas de validação e discussão dos resultados.

Esses formatos podem ser adaptados conforme o contexto, garantindo que as mensagens-chave cheguem de forma efetiva aos tomadores de decisão e demais partes interessadas. Sempre que possível, recomenda-se envolver os gestores desde o início do processo de comunicação, promovendo o entendimento mútuo dos resultados e a pactuação de encaminhamentos. A Avaliação Executiva só cumpre seu propósito se os resultados forem efetivamente utilizados: evidências geram impacto quando são comunicadas de maneira oportuna, clara e relevante, com produtos práticos que apoiem decisões e contribuam para o aprimoramento das políticas públicas.

6.4 Estrutura recomendada para o relatório de avaliação executiva

Organizar o relatório de forma clara e objetiva é fundamental para garantir a utilidade da Avaliação Executiva. A seguir, apresenta-se uma estrutura recomendada, que pode ser adaptada conforme o escopo e as necessidades da avaliação.

6.4.1 Caracterização da política avaliada

Esta seção deve apresentar o contexto, os objetivos, o público-alvo e o estágio de implementação da política. Também é importante descrever sua origem, a vinculação institucional, as bases legais e a Teoria da Mudança associada — detalhando os insumos, atividades, produtos, resultados e impactos esperados, bem como riscos e hipóteses.

Caso a Teoria da Mudança ainda não esteja formalizada, recomenda-se construí-la neste momento, para explicitar a lógica causal da intervenção.

A caracterização da política oferece ao leitor uma compreensão clara sobre o que está sendo avaliado, quais resultados se pretendem alcançar e como se espera que as mudanças ocorram.

6.4.2 Perguntas avaliativas e critérios de análise

As perguntas avaliativas devem refletir a finalidade da avaliação e cobrir diferentes pontos da cadeia lógica da política, desde a pertinência do problema até os sinais iniciais de resultados e a eficiência na execução. Em avaliações executivas, busca-se uma visão panorâmica: o problema ainda é relevante? O desenho da política é coerente e possui respaldo na literatura? A implementação está ocorrendo como previsto? Há indícios de efeitos? Os recursos estão sendo bem utilizados? Essas perguntas podem ser enriquecidas com diferentes lentes analíticas, como os critérios da OCDE (eficácia, eficiência, sustentabilidade, coerência) e abordagens transversais como equidade, gênero e mudança transformadora, permitindo uma análise mais completa da qualidade da política e da natureza dos seus efeitos.

6.4.3 Métodos e fontes de dados

Esta seção descreve o desenho metodológico adotado, os tipos de dados mobilizados e as razões para a escolha dessas abordagens. Deve esclarecer se foram utilizados dados secundários, coletas primárias ou uma combinação de fontes, indicando como os dados foram analisados e, sempre que possível, como foi feita a triangulação entre diferentes fontes e métodos para aumentar a robustez dos achados. É importante também apontar eventuais limitações, como lacunas de dados, barreiras de acesso a informantes ou restrições de tempo, e descrever as estratégias adotadas para lidar com essas fragilidades. Ao explicitar as escolhas metodológicas, a forma de uso dos dados e as estratégias de análise, esta seção reforça a credibilidade da avaliação e a consistência das interpretações apresentadas.

6.4.4 Achados e análise por dimensão

Esta seção constitui o núcleo do relatório, apresentando os achados da avaliação de forma estruturada e interpretada. Para facilitar a leitura e o vínculo direto com as perguntas avaliativas, recomenda-se organizar os resultados da avaliação por dimensões, que podem variar conforme a natureza da política. A seguir, descrevem-se subdimensões comumente utilizadas em avaliações executivas:

- **Análise do desenho e da Teoria da Mudança**: examina se a concepção da política é adequada ao problema identificado, com foco na lógica causal proposta e, sobretudo, nos riscos e nas hipóteses que ameaçam e sustentam os elos entre insumos, atividades, produtos, resultados e impactos. Trata-se de uma análise crítica da Teoria da Mudança e do desenho da intervenção,

que em geral contempla aspectos semelhantes aos critérios de relevância e coerência institucional definidos pela OCDE.

- **Implementação e processos operacionais**: descreve como a política foi executada na prática, comparando o planejado com o realizado, e identificando facilitadores, barreiras e soluções adotadas. Inclui a análise das entregas realizadas — produtos e serviços—, com base em registros administrativos e dados de monitoramento, avaliando se as ações previstas foram efetivamente executadas em termos de quantidade, qualidade e regularidade. Também são considerados aspectos como cobertura atingida, funcionamento dos processos administrativos e logísticos, engajamento das equipes e desafios enfrentados no cotidiano da implementação. As percepções de gestores, técnicos e beneficiários contribuem para contextualizar o funcionamento da política.

- **Resultados iniciais e efeitos percebidos**: descreve os sinais de mudança observados no curto e médio prazo, com base em indicadores de resultado (como frequência escolar, satisfação dos usuários ou percepção de qualidade dos serviços) e nas experiências relatadas pelos diferentes públicos. Sempre que possível, podem ser incluídas comparações entre períodos ou grupos, com a devida cautela em relação à atribuição causal.

- **Eficiência**: analisa, quando pertinente, a relação entre recursos mobilizados e resultados alcançados. Em avaliações executivas, não se calcula custo-benefício ou custo-efetividade no sentido estrito, pois não há avaliação causal de impacto. No entanto, é possível calcular indicadores simples, como custo por beneficiário ou custo por produto entregue, conforme apresentado no capítulo anterior, e apontar oportunidades de otimização.

- **Análise das lentes analíticas escolhidas**: examina dimensões transversais incorporadas à avaliação, como equidade, gênero ou sustentabilidade. Avalia se a política promove inclusão, reduz desigualdades ou gera efeitos diferenciados entre grupos sociais, analisando, por exemplo, a distribuição territorial da cobertura ou a participação de públicos historicamente excluídos.

A apresentação dos resultados pode ser enriquecida com tabelas, gráficos, boxes temáticos ou trechos de depoimentos ilustrativos. Também é recomendável utilizar quadros-resumo que destaquem os principais achados por subdimensão ou exemplos de estudos de caso, para facilitar a leitura e apoiar a interpretação dos resultados pelos gestores.

6.4.5 Conclusões e recomendações

A seção final do relatório retoma as perguntas avaliativas e apresenta, de forma sintética, as principais conclusões da avaliação, seguidas de recomendações práticas e acionáveis. As conclusões devem estar diretamente fundamentadas nos acha-

dos e oferecer respostas claras às perguntas centrais da avaliação, articulando os resultados observados com os objetivos da política.

Na sequência, são apresentadas recomendações específicas voltadas à gestão e às partes interessadas. Essas recomendações devem ser claras, viáveis e focadas em ações de curto prazo, organizadas por grau de prioridade para destacar o que é mais crítico implementar. Devem também estar diretamente baseadas nos achados da avaliação, considerar os recursos, prazos e capacidades institucionais disponíveis, e, sempre que possível, indicar os órgãos ou setores responsáveis pelos encaminhamentos.

Boas recomendações em avaliações executivas têm por objetivo apoiar decisões de forma ágil e útil, gerando impacto no curto prazo e contribuindo para o aprimoramento das políticas públicas. Ao traduzirem evidências em orientações concretas para a ação, fortalecem o uso efetivo da avaliação e o aprendizado na gestão.

Além das ações imediatas, a seção de recomendações pode também indicar temas que merecem aprofundamento futuro. Em alguns casos, os achados da Avaliação Executiva sinalizam a necessidade de avaliações mais robustas — como uma avaliação de impacto para estimar efeitos causais ou uma análise detalhada de processos para compreender melhor os gargalos na implementação. Nesses casos, a Avaliação Executiva atua como ponto de partida para uma estratégia avaliativa escalonada, ajudando a priorizar investimentos futuros em monitoramento e avaliação.

6.4.6 Anexos

Os anexos podem incluir instrumentos de coleta utilizados, tabelas ou figuras adicionais, documentos analisados, registros de oficinas ou protocolos de entrevista. Eles contribuem para reforçar a transparência da avaliação e apoiar leitores que desejam compreender detalhes metodológicos ou replicar a abordagem.

Aplicando na prática

Boas práticas para uma Avaliação Executiva

- Definir claramente o escopo e as perguntas-chave da avaliação, focando no que é mais estratégico para a tomada de decisão.
- Priorizar o uso de dados já disponíveis, combinando registros administrativos, documentos e informações de monitoramento sempre que possível.
- Envolver gestores e equipes técnicas desde o início, para garantir alinhamento de expectativas, agilizar a coleta de dados e ampliar a utilidade dos achados.
- Triangular métodos de forma ágil, combinando evidên-

cias quantitativas e qualitativas para aumentar a robustez das conclusões.

- Comunicar os achados de forma clara, objetiva e prática, com foco em entregas que ajudem na ação — como briefings executivos, apresentações e notas técnicas.

Armadilhas comuns

- Tentar responder a muitas perguntas em pouco tempo, diluindo a qualidade da análise e comprometendo a utilidade dos resultados.
- Planejar coletas extensas e complexas, incompatíveis com o tempo e os recursos disponíveis para a avaliação executiva.
- Excluir a perspectiva dos usuários da política pública, limitando a compreensão dos efeitos percebidos e das dificuldades na ponta.
- Basear conclusões apenas em entrevistas ou opiniões, sem buscar outras fontes que validem ou desafiem essas percepções.
- Produzir relatórios longos, técnicos demais e pouco úteis, que dificultam o uso prático das evidências pelas equipes gestoras.

Exemplo prático 31— Avaliação Executiva do Programa de Reforço da Alimentação Escolar

1. **Caracterização da política avaliada**

   O Programa de Reforço da Alimentação Escolar foi implementado em 50 escolas públicas municipais, com foco em áreas de maior vulnerabilidade social. Seu objetivo é melhorar a nutrição dos alunos por meio da oferta diária de uma refeição complementar no período da tarde, buscando aumentar a frequência escolar, melhorar a concentração em sala de aula e reduzir a evasão. A política também prevê ações educativas voltadas às famílias, com foco em alimentação saudável. A Teoria da Mudança do programa pressupõe que a melhoria da nutrição escolar levaria a maior disposição, menor absenteísmo e, no médio prazo, a melhores resultados educacionais.

2. **Perguntas avaliativas e critérios de análise**

   A avaliação buscou responder:

   - A implementação ocorreu como planejado, em termos de cobertura, qualidade nutricional e infraestrutura?
   - O programa é bem aceito pelos alunos, famílias e profissionais da escola?
   - Há indícios de impacto sobre a frequência escolar e o comportamento dos alunos?
   - As ações educativas foram realizadas e conseguiram envolver as famílias?
   - Os recursos foram utilizados de forma eficiente?
   - Existem diferenças de resultados por gênero ou território?

   As análises foram orientadas pelos critérios de relevância, coerência, eficiência e equidade, incorporando lentes analíticas de gênero, equidade social, sustentabilidade ambiental e mudança transformadora.

3. Métodos e fontes de dados

   A Avaliação Executiva foi realizada pela equipe da Secretaria Municipal de Educação, com apoio de consultoria técnica em nutrição.

   As fontes de dados incluíram:

   - Registros administrativos de frequência escolar e refeições servidas;
   - Entrevistas com diretores e nutricionistas;
   - Grupos focais com merendeiras e professores;
   - Questionários aplicados a famílias (200 respostas);
   - Visitas a cinco escolas selecionadas.

   Sempre que possível, os dados foram desagregados por gênero e tipo de território.

4. Resultados e análise por dimensão

   - Análise do desenho e da Teoria da Mudança

A concepção da política mostrou-se adequada ao problema identificado, com uma lógica causal consistente e uma cadeia de resultados bem definida. A definição do público-alvo foi clara, e a estratégia de intervenção se alinhou ao diagnóstico de insegurança alimentar nas escolas municipais.

- Implementação e processos operacionais

  O programa foi implementado com alta adesão (95% das escolas) e regularidade (92% dos dias letivos), atingindo cerca de 90% dos alunos presentes. Pequenas falhas logísticas foram identificadas, especialmente em escolas pequenas e rurais, mas medidas corretivas foram implementadas ainda durante a própria avaliação. A qualidade nutricional dos lanches foi mantida conforme planejado.

  A avaliação também identificou efeitos não previstos: algumas famílias relataram ter reduzido a alimentação oferecida em casa, considerando o lanche escolar como substituto suficiente para a refeição da tarde.

- Resultados iniciais e efeitos percebidos

  Os dados indicaram um aumento de 5 pontos percentuais na frequência escolar no ensino fundamental após a implementação do programa, com maior impacto em escolas de territórios mais vulneráveis. Professores e diretores relataram melhorias na disposição e atenção dos alunos no período da tarde.

- Eficiência

  O custo adicional do programa foi de aproximadamente R$ 200 mil para atender cerca de 5.000 alunos no semestre. Isso equivale a R$ 40 por aluno no semestre, ou cerca de R$ 0,50 por dia. Considerando a regularidade de 92% dos dias letivos, o custo médio por refeição efetivamente servida foi de aproximadamente R$ 0,55. A execução orçamentária alcançou 95% dos recursos previstos. Identificaram-se variações no custo por aluno entre escolas maiores e menores, sugerindo espaço para otimizações logísticas futuras.

- Análise das lentes analíticas escolhidas

  A lente da equidade indicou que o programa teve maior efeito positivo nas escolas situadas em áreas vulneráveis, contribuindo para reduzir disparidades de acesso e permanência.

  A análise de gênero mostrou que meninas e meninos aderiram igualmente ao lanche e se beneficiaram da inicia-

tiva; no entanto, o envolvimento familiar nas ações educativas concentrou-se nas mães, evidenciando padrões de desigualdade no cuidado infantil.

A dimensão da mudança transformadora apresentou avanços limitados: apesar de melhorias operacionais, ainda não se observaram alterações estruturais na gestão alimentar escolar ou na integração intersetorial.

Em relação à sustentabilidade ambiental, observou-se que a expansão do fornecimento de refeições aumentou o volume de resíduos orgânicos e embalagens. No entanto, não foram identificadas estratégias sistematizadas de redução ou reaproveitamento desses resíduos, aspecto que poderia ser incorporado em avaliações e ações futuras.

5. Conclusões e recomendações

A Avaliação Executiva concluiu que o Programa de Reforço da Alimentação Escolar foi amplamente bem-sucedido na implementação e já apresenta sinais positivos de impacto sobre a frequência escolar. A aceitação entre os beneficiários foi alta, e os ajustes necessários são pontuais e viáveis.

Com base nos achados, recomenda-se:

- Fortalecer a logística e infraestrutura, especialmente em escolas pequenas e rurais;
- Reforçar e diversificar as estratégias de comunicação e educação com as famílias, deixando claro que a refeição escolar deve complementar, e não substituir, a alimentação em casa;
- Manter o monitoramento regular da frequência escolar e planejar avaliações mais robustas para analisar efeitos de médio e longo prazo, como avaliação de impacto;
- Avaliar a possibilidade de expansão gradual do programa para outras etapas de ensino;
- Sistematizar e compartilhar boas práticas de implementação e gestão nutricional dentro e fora da rede municipal;
- Incorporar práticas de gestão ambiental, com foco em redução de desperdício alimentar e gestão de resíduos gerados pelo programa.

6. Anexos

- Roteiros de entrevistas e grupos focais utilizados na avaliação;
- Questionário aplicado às famílias;
- Tabela de cobertura e regularidade por escola;
- Estimativas de custo por aluno e custo por refeição.

Fonte: FGV CLEAR, elaboração própria.

7 Conclusão

A avaliação ex post representa uma oportunidade estratégica para fortalecer a gestão pública com base em evidências. Ao longo desta Parte III, apresentamos um conjunto integrado de abordagens — monitoramento, avaliação de implementação e processos, avaliação de resultados, avaliação de impacto, avaliação de custo efetividade e custo-benefício e avaliação executiva — que permite acompanhar, aprender e aprimorar continuamente as políticas públicas e programas sociais.

Em conjunto, essas estratégias fortalecem a capacidade da gestão pública de atuar com proatividade, foco em resultados e compromisso com a melhoria contínua e o impacto social.

A escolha da modalidade avaliativa depende do estágio de desenvolvimento da política pública, da natureza da pergunta avaliativa e do tipo de decisão que se busca informar. Avaliações ex post oferecem instrumentos valiosos para apoiar decisões qualificadas e ampliar o alcance e a efetividade das políticas.

Elas podem ser utilizadas para monitorar a execução, entender dinâmicas operacionais, verificar a obtenção de resultados, atribuir impactos, analisar a relação entre custos e benefícios ou traçar um panorama rápido para responder a demandas urgentes da gestão.

Ao integrar práticas de avaliação de forma sistemática ao ciclo da política pública, os gestores públicos reforçam a capacidade de promover soluções mais efetivas, inclusivas e sustentáveis, consolidando a cultura do uso de evidências em benefício do desenvolvimento social e institucional.

PARTE IV

Tomada de decisão

Parte IV – Tomada de decisão[23]

[23] A Parte IV deste guia se baseia na literatura técnica produzida por Al Noman, et al. (2024), Baltussen e Niessen (2006), Bamberger e Hewitt (1986), Comissão Europeia (2022), Evergreen (2014), Gertler, et al. (2016), Glewwe e Todd (2022), JRC (2020), Mackay (2007), Manski (2013), Neuman (2014), Kraft e Furlong (2019), Kusek e Rist (2004), OCDE (2020, 2022, 2025), Parkhurst (2017), Scarmo e Meagher (2017), Stetson (2008), Watkins, Meiers e Visser (2012).

Ao longo dos capítulos anteriores, este guia mostrou como formular e conduzir políticas públicas baseadas em evidências, utilizando informações robustas geradas por processos de monitoramento e avaliação para orientar decisões em cada etapa do ciclo da política pública. Na fase final desse ciclo, todas essas informações devem apoiar o gestor público na definição dos rumos futuros da política. Mas, afinal, como as evidências devem ser usadas nesse momento decisivo? Seriam elas um veredito definitivo sobre a política avaliada?

A mensagem principal desta seção é que a evidência deve auxiliar o gestor, ampliando o conjunto de informações disponíveis sobre uma situação ou contexto. Uma única evidência, porém, não deve ser interpretada como um veredito decisivo sobre a política, por diversas razões que exploraremos ao longo do capítulo. As evidências funcionam como uma bússola – podem apontar a direção para o tomador de decisão, reduzindo a incerteza e trazendo mais clareza sobre quais caminhos seguir para atingir determinado objetivo. No entanto, quem decide o caminho a tomar é o gestor (e não a bússola); por isso, o uso de evidências deve fazer parte de um processo de análise crítica e reflexiva de todas as informações disponíveis, e não ser tomado como resposta automática aos problemas públicos enfrentados.

1 A Importância da Geração e Uso de Evidências

Boas evidências geradas por meio de monitoramento e avaliação têm como objetivo retratar, com imparcialidade, a realidade de uma política pública em um dado momento e contexto. Quando produzidas com rigor, essas evidências permitem que os gestores fundamentem suas decisões em dados confiáveis, identifiquem falhas, ajustem estratégias e aprendam com experiências anteriores. Esse processo contínuo de aprendizado e adaptação é fundamental para aprimorar políticas públicas ao longo do tempo, aumentando as chances de alcançar os objetivos desejados e maximizando impactos positivos para a sociedade.

Outra contribuição valiosa do uso sistemático de evidências é a redução da assimetria de informação entre gestores e demais atores do processo decisório. Evidências estabelecem uma base comum de conhecimento, facilitando o diálogo entre governo, comunidade e especialistas, e construindo consensos por meio de análises objetivas e transparentes. Em suma, incorporar evidências nas decisões melhora a qualidade das políticas ao tornar o processo mais informado, transparente e responsável.

Entretanto, mesmo com todos esses benefícios, usar evidências na tomada de decisão não é trivial. É preciso planejamento e cultura institucional para que os dados mais adequados estejam disponíveis no momento certo. Por isso, ao longo do guia enfatizamos a geração de evidências de qualidade em cada etapa do ciclo da política, justamente para que, ao chegar na fase de decisão, o gestor tenha em mãos informações confiáveis e relevantes. Planejar antecipadamente a coleta e análise de dados garante evidências oportunas e úteis para fundamentar as decisões quando necessário. Assim, a seguir discutimos como avaliar a qualidade das evidências disponíveis e como integrá-las ao processo decisório de forma efetiva.

2 A Qualidade das Evidências

No contexto de políticas públicas, "evidência" é um termo genérico que sugere algo evidente e incontestável. No uso cotidiano, fala-se em evidência como uma prova irrefutável; porém, na prática de avaliação de políticas, existem diferentes tipos de estudos e informações que se qualificam como evidências, cada qual com seu grau de confiança e limitações. É crucial que o gestor consiga distinguir esses tipos e avaliar sua qualidade.

Um ponto de partida fundamental é verificar a qualidade dos dados brutos utilizados nas análises. Dados – quantitativos ou qualitativos – são a matéria-prima das evidências. Se os dados forem de baixa qualidade ou não representarem fielmente a realidade, qualquer avaliação realizada estará sob suspeita. Dados de baixa qualidade podem levar a conclusões inválidas sobre a política. Por exemplo, um monitoramento baseado em dados imprecisos pode indicar erroneamente que uma política está no caminho certo (quando não está) ou vice-versa. Portanto, o primeiro passo é assegurar que os indicadores e informações coletadas sejam fidedignos e robustos. Isso envolve checar metodologias de coleta, fontes de dados e consistência das medidas ao longo do tempo.

O segundo aspecto é a qualidade do estudo avaliativo que gerou a evidência. Diferentes desenhos metodológicos oferecem graus variados de confiança. Fatores como técnicas de amostragem, tamanho da amostra, métodos de coleta e estratégia de análise influenciam diretamente a credibilidade dos resultados. Estudos que buscam inferir causalidade (ou seja, atribuir mudanças observadas aos efeitos da política) exigem atenção redobrada. Se o método não for corretamente aplicado, um estudo pode indicar falsamente que "a política não teve impacto" (quando na verdade teve) ou o contrário. Mesmo com metodologias rigorosas – como experimentos controlados ou quase-experimentos – sempre haverá algum grau de incerteza estatística envolvido na medição de impacto, e os gestores devem estar cientes dessas limitações (por exemplo, por meio de intervalos de confiança e análises de sensibilidade).

Inferir causalidade em políticas públicas é um exercício complexo. Um achado pode ser válido para uma determinada amostra ou contexto, mas não ser automaticamente generalizável para outras populações. Por exemplo, um programa educacional que funcionou em uma pequena cidade pode não ter o mesmo efeito em uma metrópole, devido a diferenças socioeconômicas e institucionais. Da mesma forma, instrumentos de coleta mal desenhados podem enviesar resultados – em pesquisas do tipo survey, por exemplo, a maneira de formular uma pergunta pode influenciar a resposta do entrevistado. Assim, as fontes potenciais de viés são diversas, e identificá-las integralmente pode ser um grande desafio. No campo da saúde, conforme mencionamos ao longo deste guia, costuma-se recorrer à "pirâmide de evidências", uma hierarquia que classifica tipos de estudo conforme a con-

fiabilidade de suas evidências, colocando revisões sistemáticas e meta-análises no topo (por sintetizarem múltiplos estudos robustos), e estudos observacionais ou relatos de caso na base (por estarem mais sujeitos a vieses). Ferramentas como essa podem ser úteis em outras áreas, ajudando o gestor a julgar quão forte é o suporte empírico para determinada intervenção.

Em síntese, antes de utilizar uma evidência para decidir o futuro de uma política, deve-se perguntar: essa evidência é confiável? Avalie se os dados são acurados e pertinentes; verifique se o estudo foi conduzido com rigor metodológico; considere se os achados resistem a escrutínio estatístico e se fazem sentido em contextos análogos. Ao adotar essa postura crítica, o gestor evita tomar decisões precipitadas com base em informações possivelmente enganosas. Manski (2013) adverte que muitas análises de políticas transmitem uma certeza exagerada – às vezes chamada de "certeza incrível" – ao ignorarem ou subestimarem as incertezas inerentes aos dados e modelos utilizados. Pressupostos fortes demais nos estudos podem levar a previsões aparentemente exatas, porém pouco confiáveis, induzindo os decisores a um excesso de confiança. A alternativa proposta é explicitar a incerteza e trabalhar com o reconhecimento do conhecimento parcial: assim, os formuladores de políticas podem tomar decisões razoáveis baseadas em evidências, porém com plena consciência dos limites do que se sabe. Ou seja, evidencia-se a necessidade de humildade científica: mesmo a melhor evidência não elimina completamente a incerteza, e decisões responsáveis devem considerar tanto os dados disponíveis quanto a possibilidade de erro ou resultados imprevistos.

3 Valores, raciocínio e evidências nas tomadas de decisão

Mesmo com evidências de alta qualidade em mãos, elas por si sós não decidem – quem decide é o tomador de decisão, guiado por objetivos, valores e pelo contexto em que está inserido. A formulação, implementação e reformulação de uma política pública envolvem escolhas que vão além dos dados técnicos: entram em jogo considerações éticas, políticas e gerenciais. Assim, o uso de evidências deve ser equilibrado com o raciocínio prático e o julgamento de valor do gestor. A etapa de decisão consiste em avaliar alternativas à luz das evidências e de critérios como viabilidade política, aceitação pública e conformidade com as prioridades da sociedade. Em outras palavras, decidir implica integrar fatos e valores. A seguir, discutimos dois elementos cruciais nessa integração: a necessidade de um modelo teórico (mental) adequado para interpretar as evidências, e o reconhecimento das incertezas e critérios diversos que permeiam decisões complexas.

3.1 A Importância do modelo teórico (mental) adequado

Toda política pública carrega implícita (ou explicitamente) uma teoria de como o mundo funciona, ou seja, um modelo mental que liga as ações aos resultados esperados. Essa teoria orienta tanto o desenho da intervenção quanto a interpretação das evidências sobre seu desempenho. Ter um modelo teórico adequado é fundamental para dar sentido às evidências e tomar decisões coerentes. Evidências não "falam" por conta própria; elas precisam ser interpretadas à luz de hipóteses sobre causa e efeito. Se o gestor adota uma hipótese equivocada sobre o problema ou sobre a solução, pode acabar interpretando de maneira inadequada os dados e tomando decisões erradas.

Por exemplo, imagine uma situação de alto absenteísmo escolar em determinada comunidade. O gestor formula a teoria de que os alunos faltam às aulas por desinteresse, talvez causado por aulas pouco atrativas ou professores ausentes – problemas pedagógicos conhecidos. Com base nisso, decide investir em formação de professores e novas metodologias de ensino. No entanto, descobre-se que a principal causa das ausências era, na verdade, a dificuldade de acesso à escola: uma ponte danificada impedia a chegada dos estudantes após chuvas. Essa revelação evidencia que o modelo mental inicial estava errado – o problema não era pedagógico, e sim de infraestrutura de transporte. Se o gestor persistisse na teoria original, as evidências de aumento de faltas poderiam levá-lo a concluir, erroneamente, que as iniciativas pedagógicas não funcionaram, quando na verdade o obstáculo estava

fora da escola. Ajustar o modelo teórico (neste caso, acionando o órgão de obras para consertar a ponte) foi essencial para realinhar a solução ao verdadeiro problema. Modelos distintos levam a desenhos distintos — e, portanto, a resultados distintos.

A fim de minimizar riscos como o ilustrado, é fundamental que o modelo teórico subjacente à política seja construído com base na análise crítica de modelos prévios e no uso estratégico das evidências disponíveis. A etapa de avaliação ex ante — especialmente por meio da avaliação de desenho e da elaboração da Teoria da Mudança — é um momento oportuno para a análise e o escrutínio do modelo teórico utilizado.

Além disso, é importante reconhecer que toda interpretação de evidências também se dá por meio de um modelo mental. Os dados não falam por si: seu significado depende das suposições teóricas de quem os interpreta. Por isso, ser transparente e explícito quanto ao modelo utilizado é parte essencial do processo de tomada de decisão, permitindo maior rigor analítico, melhor comunicação com diferentes públicos e maior legitimidade das escolhas feitas.

Gestores devem estar abertos a revisar suas premissas diante de novas evidências. Na prática avaliativa, ferramentas como o Modelo Lógico e a Teoria da Mudança (TdM) formalizam esse modelo mental, detalhando insumos, atividades, produtos, resultados e impactos esperados de uma intervenção. Em suma, um modelo lógico adequado funciona como lente para enxergar o significado das evidências. Sem essa lente, corre-se o risco de interpretar dados de forma desconexa ou atribuir significado indevido a resultados isolados. Antes de decidir manter ou mudar uma política, verifique se você e sua equipe compreendem por que a política funcionou ou não funcionou – isto é, qual mecanismo foi confirmado ou refutado pelas evidências. Essa compreensão é chave para tomar decisões acertadas sobre ajustes ou expansões.

3.2 Evidências, Incertezas e Critérios de Decisão

Tomar decisões em políticas públicas exige lidar com incertezas. Mesmo diante de evidências robustas produzidas por avaliações anteriores, raramente o gestor terá segurança total sobre os resultados que se concretizarão no futuro. As evidências indicam tendências, probabilidades e efeitos observados em determinados contextos — mas não eliminam o risco inerente à ação pública. O desafio está justamente em decidir com base em conhecimento parcial, antecipando consequências em ambientes que podem mudar ao longo do tempo.

Além disso, diferentes interpretações e critérios podem levar a decisões distintas, mesmo diante do mesmo conjunto de dados. A evidência é sempre contextual: depende do momento, da população analisada, do tipo de implementação, do desenho da política e do ambiente em que ela opera. Por isso, as evidências precisam ser interpretadas com cautela e combinadas com outros elementos da decisão pública, como custo, equidade, viabilidade política, urgência e sustentabilidade.

Esses critérios, por sua vez, envolvem escolhas normativas e valores que devem ser reconhecidos e explicitados. Afinal, boas decisões públicas não se baseiam apenas na evidência disponível, mas também em julgamentos sobre o que é justo, adequado e desejável em cada realidade.

Essa multiplicidade de fatores exige que o gestor formule hipóteses sobre o futuro, pondere riscos e selecione critérios para tomar suas decisões. A teoria da decisão em contextos de incerteza ajuda a explicitar essas escolhas. Como veremos a seguir, há diferentes formas de raciocinar com base em evidências: é possível adotar critérios que maximizam o valor esperado, que evitam o pior cenário ou que minimizam o arrependimento futuro. Cada abordagem traduz um tipo de julgamento sobre o que é uma boa decisão em contextos incertos.

Box 28 – Decisões diferentes com as mesmas evidências: um exemplo sob incerteza

Considere uma gestora pública que precisa decidir entre manter um programa já existente (status quo) ou adotar uma versão inovadora. Ambos os modelos foram avaliados e apresentaram resultados positivos, mas com magnitudes distintas dependendo do contexto. Suponha que haja dois cenários possíveis no futuro (A e B) e que a taxa de sucesso de cada programa em cada contexto seja a seguinte:

Contexto	Status quo	Inovação
A	50%	90%
B	70%	40%

Como a decisão precisa ser tomada antes de saber qual contexto se realizará, a gestora precisa lidar com incerteza. Dependendo do critério de decisão adotado, a escolha pode variar.

Se a gestora estiver disposta a levar em conta a sua crença sobre as chances de realizações futuras dos cenários A ou B, ela pode utilizar o critério da regra Bayes.

- **Valor Esperado (Regra de Bayes)**

  Se a gestora acredita que o cenário A (B) ocorrerá, ela escolherá a inovação (status quo). Por sua vez, se a gestora acredita que os contextos A e B são igualmente prováveis pode calcular a média ponderada das taxas de sucesso.

  → Status quo: (50% + 70%) / 2 = 60%

  → Inovação: (90% + 40%) / 2 = 65%

  Decisão: escolher a inovação, por apresentar maior sucesso esperado.

Se a gestora não desejar levar em conta suas probabilidades subjetivas sobre as realizações futuras dos cenários A ou B, ela pode utilizar o critério MaxMin ou o critério do menor arrependimento.

- **Critério MaxMin (pior caso)**:

Se ela for extremamente avessa ao risco, pode comparar os piores resultados de cada programa entre os cenários possíveis.

→ Status quo: mínimo de 50

→ Inovação: mínimo de 40

Decisão: manter o status quo, que garante melhor desempenho no pior cenário.

- **Critério do Menor Arrependimento (Minimax Regret)**:

  Aqui, a gestora considera a perda que a sua escolha pode ter em caso desta não ser a melhor alternativa.

  → Se escolher status quo, o arrependimento no status quo para cada cenário é:

  - No A: 90% – 50% = 40%
  - No B: 70% – 70% = 0%

Assim, o máximo arrependimento no status quo é 40%.

→ Se escolher a inovação, o arrependimento na inovação para cada cenário é:

- No A: 90% – 90% = 0%
- No B: 70% – 40% = 30%

Decisão: Seja δ a fração da população que recebe o modelo de inovação. Se o novo modelo for pior que o do status quo, o máximo arrependimento é 30% × δ. Se o novo modelo com inovação for melhor que o status quo, o máximo arrependimento é 40% × (1 – δ).

O máximo arrependimento é minimizado quando:

$$30\% \times \delta = 40\% \times (1 - \delta)$$

$$\delta = \frac{40}{70} \cong 0{,}57$$

Ou seja, expande-se o modelo de inovação para 57% da população e o status quo para os 43% restantes.

Esse exemplo mostra que não há uma única decisão correta com base nos dados. Tudo depende de como o tomador de decisão lida com incertezas e quais valores prioriza — segurança, inovação, equilíbrio de riscos, entre outros. Evidências são insumos essenciais, mas a escolha entre caminhos possíveis envolve também julgamento e critérios explícitos de decisão.

Fonte: FGV CLEAR, elaboração própria.

3.2.1 Decidir na prática: evidências como guia, não como sentença

O exemplo acima ilustra como diferentes critérios de decisão — todos legítimos — podem levar a escolhas distintas a partir do mesmo conjunto de evidências. Na prática da gestão pública, essas decisões são ainda mais complexas: além da incerteza sobre os contextos futuros, o gestor precisa considerar variáveis como restrições orçamentárias, viabilidade política, urgência do problema, participação de atores relevantes e impactos sobre grupos específicos da população.

Por isso, avaliações e outras fontes de evidência devem ser entendidas como **instrumentos para iluminar possibilidades**, e não como vereditos automáticos sobre o que deve ser feito. A melhor decisão não é necessariamente manter o que funcionou ou eliminar o que não funcionou. Muitas vezes, os dados indicam a necessidade de ajustes, redirecionamentos ou mesmo de estratégias completamente novas. O papel do gestor é **combinar a evidência empírica com critérios técnicos, éticos e políticos**, construindo decisões equilibradas e justificadas.

3.2.2 Critérios práticos que influenciam a decisão

Além dos critérios formais de decisão, como maximização do valor esperado ou minimização do arrependimento, os gestores públicos enfrentam variáveis adicionais que afetam — e muitas vezes definem — a decisão final. Evidências robustas sobre impacto não eliminam a necessidade de considerar questões como:

- **Custos e recursos disponíveis**: Uma política pode ser efetiva, mas seu custo pode inviabilizar a continuidade ou expansão. É necessário avaliar se há orçamento disponível e se o custo por benefício é aceitável em comparação com alternativas. Ferramentas como a análise de custo-benefício ou custo-efetividade são úteis nessa análise.
- **Equidade e impactos distributivos**: Uma média de impacto positiva pode esconder desigualdades. Políticas que não alcançam grupos vulneráveis podem demandar ajustes, mesmo que sejam efetivas em termos agregados. Garantir que os benefícios sejam bem distribuídos é parte de uma gestão pública justa.
- **Viabilidade política e legal**: Mesmo que a evidência recomende encerrar uma política ineficaz, essa decisão pode enfrentar resistência de beneficiários, gestores intermediários ou legisladores. O gestor precisa avaliar o ambiente político e a janela de oportunidade para mudança.

- **Urgência e tolerância a risco**: Em cenários de emergência, como crises sanitárias ou desastres naturais, pode ser necessário agir com base em evidências ainda preliminares. Já em políticas de longo prazo, é prudente esperar por avaliações mais completas antes de expandir ou reformular.
- **Sustentabilidade e alinhamento estratégico**: Mudanças no contexto econômico, demográfico ou político podem exigir redirecionamento de políticas — mesmo daquelas que funcionaram no passado. Monitoramento contínuo é essencial para captar sinais de que o cenário mudou.

3.2.3 Caminhos possíveis após a avaliação: manter, expandir, ajustar ou encerrar

Ao final de uma avaliação, diferentes decisões podem ser tomadas. A evidência disponível aponta direções, mas a decisão final será sempre o resultado de uma síntese entre os achados empíricos, o julgamento técnico e político do gestor e as condições do contexto. Abaixo, detalhamos os quatro caminhos principais e os elementos que devem ser considerados em cada um deles:

- **Manter a política**:

  Quando a intervenção atinge seus objetivos de forma satisfatória e continua relevante, a decisão pode ser manter sua implementação tal como está. Evidências que confirmam o alcance das metas previstas (por exemplo, melhoria de aprendizagem em um programa educacional) fornecem suporte para sua continuidade. Ainda assim, políticas bem-sucedidas devem ser monitoradas continuamente, para garantir que as condições que favoreceram seu sucesso (financiamento, equipe, adesão do público-alvo) permaneçam. "Não mexer em time que está ganhando" não significa estagnação — sempre há espaço para melhorias incrementais.

- **Expandir ou replicar**:

  Resultados acima das expectativas podem justificar a ampliação da política para novos públicos, regiões ou escalas. No entanto, é necessário avaliar se o sucesso observado se deve a características replicáveis (como o desenho da política) ou a condições específicas do piloto. Muitas vezes, é prudente expandir gradualmente, mantendo a avaliação ativa durante o processo. Isso permite aprender com a expansão e ajustar em tempo real.

- **Ajustar ou reformular**:

  Nem sempre um resultado aquém do esperado significa fracasso total. Avaliações podem revelar aspectos do desenho ou da implementação que precisam ser corrigidos. Pode-se, por exemplo, simplificar processos, reforçar componentes mais promissores ou redefinir o público-alvo. Ajustar com base em evidências é sinal de aprendizado e responsabilidade — e o ciclo ideal envolve nova avaliação após os ajustes para verificar sua efetividade.

- **Encerrar ou substituir**:

  Quando os dados mostram de forma consistente que a política não gera os resultados desejados — ou mesmo produz efeitos adversos — a decisão mais responsável pode ser a descontinuidade. Encerrar um programa público é sempre delicado, mas insistir em políticas ineficazes gera custos sociais e orçamentários. Nesse caso, é recomendável comunicar os motivos de forma transparente e considerar alternativas com base em experiências bem-sucedidas em contextos semelhantes.

Em qualquer dos cenários — manter, expandir, ajustar ou encerrar — as evidências atuam como base da decisão, informando o gestor sobre o que funcionou, em que medida e por quê. Resultados de avaliações ex post subsidiam diretamente esse processo: quando estão de acordo ou acima das metas previstas, sinalizam manutenção ou expansão; quando ficam aquém, sugerem ajustes ou descontinuidade.

Contudo, é fundamental destacar que a decisão final envolve mais do que os dados: ela exige ponderar as evidências à luz dos critérios mencionados (como custos, equidade e viabilidade política) e do julgamento do gestor sobre o contexto. Pode-se, por exemplo, optar por expandir um programa com impacto moderado se ele atender a uma demanda social relevante e não houver alternativa mais eficaz disponível — trata-se de um julgamento de valor legítimo. Da mesma forma, pode-se decidir manter um programa piloto por mais tempo antes de escalar, com o objetivo de acumular evidências adicionais e reduzir incertezas.

Assim, o uso produtivo das evidências na tomada de decisão requer tanto consciência técnica (sobre o grau de confiança que os dados oferecem) quanto sabedoria prática (para ponderar cuidadosamente interesses, riscos e consequências). No próximo tópico, exploraremos como o contexto institucional e político influencia esse equilíbrio entre a evidência e outros fatores na decisão pública.

4 O Contexto Político

Nos últimos anos, observa-se um interesse crescente em políticas públicas orientadas por evidências. Iniciativas de organizações internacionais, centros de pesquisa e governos têm difundido métodos de avaliação e dados abertos para embasar decisões. No entanto, é crucial reconhecer que decisões de políticas ocorrem em arenas políticas, onde múltiplos atores – políticos eleitos, burocratas, grupos de interesse, mídia e cidadãos – disputam narrativas e preferências. Nesse ambiente, as evidências científicas são altamente desejáveis, mas não falam por si mesmas: elas serão interpretadas e, por vezes, estrategicamente utilizadas conforme as agendas em jogo.

Um mesmo conjunto de evidências pode gerar leituras distintas. Considere que uma política avaliada apresente resultados positivos em certos aspectos e negativos em outros. Grupos diferentes podem selecionar apenas a parte das evidências que apoia sua posição, proclamando "esta evidência prova que minha proposta é a melhor". Por exemplo, uma intervenção econômica pode aumentar a renda média da população (o que agrada ao formulador), mas também ampliar a desigualdade regional (o que alimenta críticas da oposição). Ambos os lados podem dizer que "a evidência está do meu lado", se focarem apenas no indicador que lhes convém. Esse cenário ressalta o risco do viés de confirmação na utilização de evidências – isto é, dar atenção apenas aos dados que confirmam crenças prévias ou interesses próprios.

Como solucionar esse dilema? Em primeiro lugar, é necessária uma ética no uso da evidência pelo setor público. Os gestores devem se comprometer a avaliar honestamente todo o conjunto de resultados, reconhecendo os pontos fortes e fracos da política. Transparência é fundamental: dados e relatórios de avaliação devem ser disponibilizados de forma aberta, permitindo que diferentes análises sejam feitas e validadas. Assim, evita-se que apenas cifras convenientes ganhem destaque. Instituições de controle e a sociedade civil podem ajudar a escrutinar o uso seletivo de informação, cobrando justificativas embasadas dos tomadores de decisão.

Adicionalmente, autores como Justin Parkhurst (2017) propõem fortalecer a "boa governança da evidência" no setor público. Esse conceito refere-se à criação de arranjos institucionais que garantam a integração de evidências rigorosas, sistemáticas e tecnicamente válidas aos processos de decisão de modo inclusivo e responsável. Na prática, isso significa construir instituições de aconselhamento ou unidades de análise dentro do governo que operem com dupla preocupação: de um lado, aderência aos princípios científicos (rigor metodológico, revisão por pares, replicabilidade dos estudos, transparência de dados); de outro, representação dos diversos interesses democráticos (ouvir atores envolvidos e afetados, considerar impactos para diferentes grupos da população).

Um exemplo são comitês consultivos com especialistas independentes que publi-

quem recomendações baseadas em evidências para determinadas políticas, ou os "laboratórios de inovação" e núcleos de avaliação em ministérios, que trazem análises imparciais ao debate interno. Tais arranjos podem reduzir a influência de pressões político-partidárias de curto prazo na interpretação das evidências, ao mesmo tempo em que tornam o processo mais aberto e plural. Parkhurst (2017) argumenta que o uso de evidências no governo só alcançará todo o seu potencial para melhorar as políticas públicas se incorporar, simultaneamente, princípios tanto da boa ciência (como objetividade e rigor) quanto da boa governança democrática (como prestação de contas e inclusividade).

Outro aspecto fundamental do contexto político é a cultura organizacional e os incentivos existentes dentro do setor público. Como destaca Mackay (2007), para que decisões baseadas em evidências se tornem uma prática sistemática, é necessário institucionalizar sistemas de monitoramento e avaliação que sejam valorizados pelos órgãos governamentais. Quando há incentivos formais — como a exigência de avaliações independentes para renovação de programas, a vinculação orçamentária a resultados comprovados ou o reconhecimento por inovações orientadas por evidências — os gestores tendem a incorporar com mais seriedade os dados em seus processos decisórios.

Esse movimento tem ganhado força em diversos países da OCDE, impulsionado pela Recomendação sobre Avaliação de Políticas Públicas (OCDE, 2022), que incentiva a integração sistemática da avaliação ao ciclo de políticas, com responsabilidades claras e estímulos institucionais para o uso dos achados.

A experiência acumulada pelo FGV CLEAR nos últimos dez anos — tanto no Brasil quanto internacionalmente, por meio da Global Evaluation Initiative — corrobora essa posição. Governos que institucionalizam sistemas de avaliação costumam avançar na construção de uma governança robusta para a geração e o uso de evidências, o que se reflete em maior uso efetivo dessas evidências na formulação e reformulação de políticas públicas. Há exemplos concretos em diversos países e no Brasil, tanto no nível federal quanto em governos subnacionais, que demonstram como a criação de estruturas, regras e incentivos adequados fortalece uma cultura de gestão pública orientada para resultados.

Por fim, é preciso reconhecer que decisões públicas nem sempre seguirão a evidência disponível à risca — e isso não significa, necessariamente, que os gestores estejam desinformados ou que falte compromisso com a boa gestão. Políticas públicas envolvem valores, disputas e escolhas coletivas. Haverá situações em que, mesmo diante de evidências de efetividade modesta, opta-se por manter um programa por razões sociais ou políticas legítimas — como, por exemplo, um programa cultural que fortalece identidades locais, embora seu impacto econômico seja limitado.

Em outros casos, as evidências podem apontar para uma direção clara, mas sua implementação exige tempo e capacidade de negociação para construir apoio político e social. Nesses contextos, cabe ao gestor público atuar como um "tradutor" das evidências, comunicando os achados de forma clara e acessível a parlamentares, imprensa e cidadãos. O objetivo é promover um debate político informado por dados concretos — e não por percepções equivocadas ou boatos. Ferramentas

como infográficos, painéis públicos de indicadores, vídeos curtos ou relatos de beneficiários ajudam a contar a história por trás dos números e a engajar diferentes públicos.

Em síntese, o contexto político influencia se, como e quando as evidências serão utilizadas. Reconhecer essa realidade evita uma visão ingênua de que "basta entregar o relatório técnico para que a boa decisão seja tomada". É necessário investir em instituições e processos que assegurem a integridade do uso da evidência, assim como em estratégias de comunicação e engajamento que insiram os achados avaliativos no centro do debate público. Dessa forma, maximiza-se a chance de que as evidências cumpram seu papel de aprimorar políticas públicas — mesmo em ambientes politicamente contestados.

5 Conclusão

Na Parte IV, exploramos como as evidências podem — e devem — apoiar a tomada de decisões sobre o futuro das políticas públicas. Fechando o ciclo proposto ao longo do guia, reiteramos que as informações geradas em cada etapa — do diagnóstico à avaliação — devem funcionar como ferramentas que orientam o gestor na definição dos rumos da política. Evidências de qualidade, produzidas com rigor e submetidas a uma análise criteriosa, são insumos essenciais para uma gestão pública mais informada, transparente e responsável.

Ao mesmo tempo, reafirmamos que o uso de evidências não se confunde com a aplicação automática de regras. Decidir envolve lidar com incertezas, considerar contextos diversos e fazer escolhas que dialogam com valores, interesses e prioridades sociais. As evidências não são oráculos infalíveis: são bússolas que orientam, mas não substituem o julgamento e a responsabilidade dos tomadores de decisão.

Discutimos também como diferentes critérios — formais e práticos — influenciam a decisão, desde modelos de raciocínio sob incerteza até fatores como custo, equidade, viabilidade política e urgência. Mostramos que, diante dos resultados de uma avaliação, o gestor pode optar por manter, expandir, ajustar ou encerrar uma política. E que qualquer uma dessas escolhas exige uma análise combinada de dados, objetivos e restrições.

Além disso, destacamos o papel central do contexto político e da governança institucional para viabilizar o uso efetivo das evidências. Decisões públicas não ocorrem em um vácuo técnico, mas em ambientes onde múltiplos atores disputam visões e interesses. Promover uma cultura organizacional que valorize a avaliação, criar incentivos e estruturas adequadas e desenvolver estratégias eficazes de comunicação são elementos indispensáveis para fortalecer o uso legítimo da evidência na gestão pública.

Por fim, reconhecemos que nem toda boa decisão seguirá à risca a melhor evidência disponível — e isso não significa, necessariamente, erro, negligência ou desinformação. Significa, muitas vezes, considerar aspectos legítimos que extrapolam os dados.

Nesse cenário, o papel do gestor é integrar evidências robustas com raciocínio crítico e sensibilidade institucional, construindo decisões que sejam não apenas informadas, mas também coerentes com os desafios concretos da sociedade. O essencial é que essas decisões sejam justificadas com racionalidade e transparência.

Referências

1 ÁFRICA DO SUL. DEPARTAMENTO DE PLANEJAMENTO, MONITORAMENTO E AVALIAÇÃO (DPME). **Evaluation Guideline No. 2.2.21: How to undertake rapid evaluations**. Pretória, 2020. Disponível em: <https://evaluations.dpme.gov.za/images/gallery/GL%202.2.21%20How%20to%20Undertake%20Rapid%20Evaluations.pdf>. Acesso em: 4 mar. 2024.

2 AL NOMAN, A. et al. Simplifying the concept of level of evidence in lay language for all aspects of learners: In brief review. **Intelligent Pharmacy**, v. 2, n. 2, p. 270–273, abr. 2024. DOI: 10.1016/j.ipha.2023.11.002. Disponível em: <https://doi.org/10.1016/j.ipha.2023.11.002>. Acesso em: 7 abr. 2025.

3 ANGRIST, Joshua D.; PISCHKE, Jörn-Steffen. **Mostly harmless econometrics: an empiricist's companion**. Princeton: Princeton University Press, 2009.

4 ASSOCIADOS DE PESQUISA E AVALIAÇÃO DO PACÍFICO (PREA); SECRETARIADO DA COMUNIDADE DO PACÍFICO (SPC). **Learner Guide: Proposal preparation using the logical framework approach & project monitoring – Part II**. Suva, 2015. Disponível em: <https://www.prea.com.au/images/toolbox/Learner_Guide_SPC_LFA_May_2015_v1.5.pdf>. Acesso em: 7 abr. 2025.

5 BALTUSSEN, R.; NIESSEN, L. Priority setting of health interventions: the need for multi-criteria decision analysis. **Cost Effectiveness and Resource Allocation**, v. 4, p. 14, ago. 2006. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC1560167/>. Acesso em: 7 abr. 2025.

6 BAMBERGER, Michael; ANGELI, J. Michael Kirk. **Making smart policy: Using impact evaluation for policy making - case studies on evaluations that influenced policy**. Washington, DC: Banco Mundial, 2009.

7 BAMBERGER, Michael; HEWITT, Eleanor. **Monitoring and evaluating urban development programs: a handbook for program managers and researchers**. Washington, DC: World Bank, 1986.

8 BAMBERGER, Michael; RUGH, Jim; MABRY, Linda. **RealWorld Evaluation: working under budget, time, data, and political constraints**. 2. ed. Thousand Oaks: SAGE Publications, 2012.

9 BANCO ASIÁTICO DE DESENVOLVIMENTO (ADB). **Cost-benefit analysis for development: a practical guide**. Mandaluyong, 2013. Disponível em: <https://www.adb.org/documents/cost-benefit-analysis-development-practical-guide>. Acesso em: 17 abr. 2025.

10 BANCO EUROPEU DE INVESTIMENTO (BEI). **The Economic Appraisal of Investment Projects at the EIB**. Luxemburgo, 2013.

11 BANCO INTERAMERICANO DE DESENVOLVIMENTO (BID). **Guía para el diseño de proyectos con enfoque de marco lógico**. Washington, DC, 2010. Disponível em: <https://publications.iadb.org/publications/spanish/document/Gu%C3%ADa-para-el-dise%C3%B1o-de-proyectos-con-enfoque-de-marco-l%C3%B3gico.pdf>. Acesso em: 1 abr. 2025.

12 _____. **Pequenos empurrões para a América Latina e o Caribe: uma década de melhoria da política pública com a ciência comportamental**. [S.l.], 2022. Disponível em: <https://publications.iadb.org/es/pequenos-empujones-para-america-latina-y-el-caribe-una-decada-de-mejorar-la-politica-publica-con-la>. Acesso em: 1 abr. 2025.

13 BANCO MUNDIAL. **Cost-Benefit Analysis in World Bank Projects**. Washington, DC, 2011.

14 _____. **Executive evaluations. Results of the Expert Roundtables on Innovative Performance Measurement Tools, Note no. 2**. Washington, DC, 2012. Disponível em: <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/617881468263940441/executive-evaluations>. Acesso em: 4 mar. 2024.

15 _____. **Public sector performance indicators**. Washington, DC, 2004.

16 _____. **World Development Report 2018: Learning to Realize Education's Promise**. Washington, DC: World Bank, 2018. Disponível em: <https://www.worldbank.org/en/publication/wdr2018>. Acesso em: 11 abr. 2025.

17 _____. **World Development Report 2019: The Changing Nature of Work**. Washington, DC: World Bank, 2019. Disponível em: <https://www.worldbank.org/en/publication/wdr2019>. Acesso em: 11 abr. 2025.

18 BETTEREVALUATION. **Footprint Evaluation**. [S.l.], 2022. Disponível em: <https://www.betterevaluation.org/methods-approaches/themes/footprint-evaluation>. Acesso em: 8 abr. 2025.

19 _____. **Problem tree analysis**. [S.l.], 2014. Disponível em: <https://www.betterevaluation.org/en/evaluation-options/problem_tree>. Acesso em: 1 abr. 2025.

20 _____. **Rainbow framework**. [S.l.], 2023. Disponível em: <https://www.betterevaluation.org/en/rainbow_framework>. Acesso em: 11 abr. 2025.

21 _____. **Risk assessment**. [S.l.], 2025. Disponível em: <https://www.betterevaluation.org/methods-approaches/methods/risk-assessment>. Acesso em: 7 abr. 2025.

22 BETTEREVALUATION. **Stakeholder mapping and analysis**. [S.l.], 2024. Disponível em: <https://www.betterevaluation.org/methods-approaches/methods/stakeholder-mapping>. Acesso em: 30 mar. 2025.

23 BOARDMAN, Anthony E. et al. **Cost-benefit analysis: concepts and practice**. 4. ed. New Jersey: Prentice Hall, 2018.

24 BRASIL. **Avaliação de Políticas Públicas: Guia Prático de Análise ex ante**. Brasília, DF, 2018. Disponível em: <https://www.gov.br/casacivil/pt-br/assuntos/downloads/153743_analise-ex-ante_web_novo.pdf/view>. Acesso em: 16 abr. 2025.

25 _____. **Avaliação de políticas públicas: guia prático de análise ex post**. Brasília, DF, 2018. Disponível em: <https://www.gov.br/casacivil/pt-br/assuntos/downloads/guiaexpost.pdf/view>. Acesso em: 16 abr. 2025.

26 BRASIL. TRIBUNAL DE CONTAS DA UNIÃO. **Referencial de controle de políticas públicas**. Brasília, 2020. Disponível em: <https://portal.tcu.gov.br/data/files/47/17/2B/CC/50D729104CE08619E18818A8/Referencial%20de%20Controle%20de%20Politicas%20Publicas.pdf>. Acesso em: 3 abr. 2025.

27 CABO OCIDENTAL. GOVERNO PROVINCIAL. **Cape Evaluation Guideline No. 4: How to undertake rapid evaluations**. Western Cape, 2020. Disponível em: <http://www.twendembele.org/reports/dpme-rapid-evaluation-guideline>. Acesso em: 4 mar. 2024.

28 _____. **Getting Evidence Quicker: The development of rapid evaluation in selected African countries**. Policy Brief. [S.l.], 2021. Disponível em: <http://twendembele.org/reports/getting-evidence-quicker/>. Acesso em: 4 mar. 2024.

29 CDC – CENTERS FOR DISEASE CONTROL AND PREVENTION. **Types of evaluation: formative, process, outcome and impact**. Atlanta, 2017. Disponível em: <https://www.cdc.gov/evaluation>. Acesso em: 11 abr. 2025.

30 CENTRO AUSTRALIANO PARA AVALIAÇÃO (ACE). **How to evaluate**. [S.l.], 2025. Disponível em: <https://evaluation.treasury.gov.au/toolkit/how-evaluate#top>. Acesso em: 28 abr. 2025.

31 CENTRO COMUM DE INVESTIGAÇÃO DA COMISSÃO EUROPEIA – JRC. **10 tips of researchers: How to achieve impact on policy?** Bruxelas, 2020. Disponível em: <https://knowledge4policy.ec.europa.eu/publication/10-tips-researchers-how-achieve-impact-policy_en>. Acesso em: 28 abr. 2025.

32 CENTRO DE APRENDIZAGEM EM AVALIAÇÃO E RESULTADOS PARA O BRASIL E ÁFRICA LUSÓFONA (FGV CLEAR). **Diagnóstico dos Sistemas de Avaliação de Políticas Públicas no Brasil**. [S.l.], 2025.

33 COLÔMBIA. DEPARTAMENTO NACIONAL DE PLANEJAMENTO. **Banco de Herramientas de la Política de Mejora Normativa**. [S.l.], 2024. Disponível em: <https://www.dnp.gov.co/LaEntidad_/subdireccion-general-prospectiva-desarrollo-nacional/direccion-gobierno-ddhh-paz/Paginas/herramientas-de-la-politica-de-mejora-normativa.aspx>. Acesso em: 9 abr. 2025.

34 COLÔMBIA. GRUPO DE MODERNIZACIÓN DEL ESTADO. **Guía Metodológica para la Elaboración de Análisis de Impacto Normativo (AIN)**. Bogotá, 2021. P. 1–77.

35 COMISSÃO EUROPEIA. **Commission staff working document: Supporting and connecting policymakers in the Member States with scientific research**. Bruxelas, 2022. Disponível em: <https://knowledge4policy.ec.europa.eu/file/staff-working-document-supporting-connecting-policymaking-member-states-scientific-research_en>. Acesso em: 28 abr. 2025.

36 _____. **Guide to Cost-Benefit Analysis of Investment Projects: Economic appraisal tool for Cohesion Policy 2014-2020**. Bruxelas, 2015. Disponível em: <https://ec.europa.eu/regional_policy/en/information/publications/guides/2014/guide-to-cost-benefit-analysis-of-investment-projects-for-cohesion-policy-2014-2020>. Acesso em: 17 abr. 2025.

37 COMISSÃO EUROPEIA. DG INTPA. **Evaluation Handbook**. [S.l.], 2024. Disponível em: <https://op.europa.eu/en/publication-detail/-/publication/06350947-4d56-11ef-acbc-01aa75ed71a1/language-en>. Acesso em: 8 abr. 2025.

38 CONEVAL – CONSEJO NACIONAL DE EVALUACIÓN DE LA POLÍTICA DE DESARROLLO SOCIAL. **Tipos de evaluaciones y sus características**. México, 2024. Disponível em: <https://chepes.sedesol.gob.hn/wp-content/uploads/2024/10/6.Tipos-de-Evaluacion.pdf>. Acesso em: 15 abr. 2025.

39 CRESWELL, J. W.; CLARK, V. L. P. **Designing and conducting mixed methods research**. Los Angeles: SAGE Publications, 2011.

40 DEPARTMENT OF PLANNING, MONITORING AND EVALUATION (DPME). **Evaluation Guideline No. 2.2.21: How to undertake rapid evaluations**. Pretória, 2020. Disponível em: <https://evaluations.dpme.gov.za/images/gallery/GL%202.2.21%20How%20to%20Undertake%20Rapid%20Evaluations.pdf>. Acesso em: 4 mar. 2024.

41 DHALIWAL, I. et al. Comparative cost-effectiveness analysis to inform policy in developing countries: a general framework with applications for education. In: GLEWWE, P. (Ed.). **Education policy in developing countries**. Chicago: University of Chicago Press, 2013. P. 285–338.

42 DNP – DEPARTAMENTO NACIONAL DE PLANEACIÓN (COLÔMBIA). **Guía metodológica de evaluación de programas públicos**. Bogotá, 2021.

43 DRUMMOND, Michael F. et al. **Methods for the economic evaluation of health care programmes**. 4. ed. Oxford: Oxford University Press, 2015.

44 EVERGREEN, Stephanie D. H. **Presenting data effectively: communicating your findings for maximum impact**. Los Angeles: SAGE, 2014.

45 FIOCRUZ; FUNDACIÓN EBERT. **Siembra y Cosecha: Manual de evaluación participativa**. [S.l.]. Disponível em: <https://evalparticipativa.net>. Acesso em: 1 abr. 2025.

46 FISHER, C.; ANUSHKO, A. Research ethics in social science. In: ALASUNTARI, P.; BICKMAN, L.; BRANNEN, J. (Ed.). **The SAGE Handbook of Social Research Methods**. London: Sage Publications, 2008.

47 FUNDO DAS NAÇÕES UNIDAS PARA A INFÂNCIA (UNICEF). **Formative evaluation handbook**. New York, 2021. Disponível em: <https://www.unicef.org/evaluation/documents/formative-evaluation-handbook>. Acesso em: 11 abr. 2025.

48 FUNDO DE POPULAÇÃO DAS NAÇÕES UNIDAS (UNFPA). **Evaluation Handbook**. [S.l.], 2024.

49 FUNDO GLOBAL PARA O MEIO AMBIENTE (GEF). **Achieving transformation through GEF investments: Information Brief**. Washington, DC, 2022. Disponível em: <https://www.thegef.org/sites/default/files/documents/2022-05/EN_GEF.STAP_.C.62.Inf_.05_Achieving_Transformation_through_GEF_Investments.pdf>. Acesso em: 9 abr. 2025.

50 FUNDO GLOBAL PARA O MEIO AMBIENTE (GEF) – ESCRITÓRIO INDEPENDENTE DE AVALIAÇÃO (GEF IEO). **Evaluation of GEF support for transformational change: evaluation report no. 122**. Washington, DC, 2018. ISBN 978-1-64233-005-2. Disponível em: <https://www.gefieo.org>. Acesso em: 8 abr. 2025.

51 FUNDOS DE INVESTIMENTO CLIMÁTICO (CIF). **Transformational change concepts: working definition, dimensions, questions and signals**. Washington, DC, 2024. Disponível em: <https://www.climateinvestmentfunds.org>. Acesso em: 8 abr. 2025.

52 FUNNELL, S.; ROGERS, P. **Purposeful program theory: effective use of theories of change and logic models**. San Francisco: Jossey-Bass, 2011.

53 GERTLER, P. J. et al. **Impact evaluation in practice**. 2. ed. Washington, DC: The World Bank, 2016. Disponível em: <https://openknowledge.worldbank.org/entities/publication/ebbe3565-69ff-5fe2-b65d-11329cf45293>. Acesso em: 8 abr. 2025.

54 GLEWWE, Paul; TODD, Petra. **Impact evaluation in international development: theory, methods, and practice**. Washington, DC: World Bank, 2022.

55 GRUPO DE AVALIAÇÃO DAS NAÇÕES UNIDAS (UNEG). **UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation**. [S.l.], 2024. Disponível em: <https://unevaluation.org>. Acesso em: 8 abr. 2025.

56 HENRIQUES, R.; CARVALHO, M.; BITTAR, M. **Gestão na educação em larga escala: Jovem de Futuro: de projeto piloto em escolas para uma política de rede pública**. São Paulo, SP: Instituto Unibanco, 2020.

57 IMAS, Linda G.; RIST, Ray C. **The road to results: designing and conducting effective development evaluations**. Washington, DC: The World Bank, 2009. Disponível em: <https://documents1.worldbank.org/curated/en/400101468169742262/pdf/The-road-to-results-designing-and-conducting-effective-development-evaluations.pdf>. Acesso em: 8 abr. 2025.

58 IMBENS, Guido W.; RUBIN, Donald B. **Causal Inference for Statistics, Social, and Biomedical Sciences: an Introduction**. Cambridge: Cambridge University Press, 2015.

59 INDEPENDENT EVALUATION GROUP (IEG). **Methods Sourcebook**. [S.l.], 2025. Disponível em: <https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/MethodsSourceBook.pdf>. Acesso em: 12 abr. 2025.

60 INSTITUTO JONES DOS SANTOS NEVES (IJSN). **Guia para Avaliar Políticas Públicas: A política é nova? Avaliação ex ante!** Vitória, 2018. Disponível em: <https://ijsn.es.gov.br/Media/IJSN/PublicacoesAnexos/cadernos/01_Volume-1_A_politica_e_nova-Avalicao_ex_ante_v3.pdf>. Acesso em: 16 abr. 2025.

61 _____. **Guia para Avaliar Políticas Públicas: Avaliação ao alcance de todos – análise executiva**. Vitória, 2018. Disponível em: <https://ijsn.es.gov.br/Media/IJSN/PublicacoesAnexos/livros/IJSN_SiMAPP_Volume-03-1.pdf>. Acesso em: 16 abr. 2025.

62 _____. **Guia para Avaliar Políticas Públicas: Como monitorar uma política pública?** Vitória, 2018. Disponível em: <https://ijsn.es.gov.br/Media/IJSN/PublicacoesAnexos/livros/IJSN_SiMAPP_Volume-02.pdf>. Acesso em: 16 abr. 2025.

63 _____. **Guia para Avaliar Políticas Públicas: E quando a política está em andamento? Avaliação ex post!** Vitória, 2018. Disponível em: <http://www.ijsn.es.gov.br/component/attachments/download/6421>. Acesso em: 20 mar. 2023.

64 INSTITUTO JONES DOS SANTOS NEVES (IJSN). **Relatório Anual de Monitoramento e Avaliação – Ciclo 2019**. Vitória, 2019. Disponível em: <https://ijsn.es.gov.br/Media/IJSN/PublicacoesAnexos/relatorios/Relatorio-SiMAPP-Ciclo-2019.pdf>. Acesso em: 4 mar. 2024.

65 KRAFT, M. E.; FURLONG, S. R. **Public policy: politics, analysis, and alternatives**. [S.l.]: CQ Press, 2019.

66 KUSEK, Jody Zall; RIST, Ray C. **Ten steps to a results-based monitoring and evaluation system: a handbook for development practitioners**. Washington, DC: The World Bank, 2004. Disponível em: <https://openknowledge.worldbank.org/server/api/core/bitstreams/f87d81cf-54e9-5a35-ab9e-dc24fc61f85a/content>. Acesso em: 8 abr. 2025.

67 LEVIN, Henry M.; MCEWAN, Patrick J. **Cost-effectiveness analysis: methods and applications**. 2. ed. Thousand Oaks: Sage Publications, 2001.

68 LIMA, L.; SOUZA, A. P. (Ed.). **Análise de custo-benefício**. São Paulo: Fundação Getulio Vargas, 2025.

69 _____. **Avaliação de Impacto**. São Paulo: Fundação Getulio Vargas, 2025.

70 _____. **Avaliação de impacto: método de Diferença em Diferenças (DiD)**. São Paulo: Fundação Getulio Vargas, 2025.

71 _____. **Avaliação ex-ante de impacto: modelagem de microssimulação**. São Paulo: Fundação Getulio Vargas, 2025.

72 _____. **Cálculo de poder estatístico**. São Paulo: Fundação Getulio Vargas, 2025.

73 MACKAY, Keith. **How to build M&E systems to support better government**. Washington, DC: World Bank, 2007. Disponível em: <https://hdl.handle.net/10986/6851>. Acesso em: 26 abr. 2025.

74 MANBY, L. et al. Healthcare workers' perceptions and attitudes towards the UK COVID-19 vaccination programme. **medRxiv**, 2021. DOI: 10.1101/2021.03.30.21254459. Disponível em: <https://doi.org/10.1101/2021.03.30.21254459>. Acesso em: 4 mar. 2024.

75 MANSKI, Charles F. **Public policy in an uncertain world: analysis and decisions**. Cambridge, MA: Harvard University Press, 2013.

76 MCNALL, Miles; FOSTER-FISHMAN, Pennie G. Methods of Rapid Evaluation, Assessment, and Appraisal. **American Journal of Evaluation**, v. 28, n. 2, p. 151–168, 2007.

77 MENEZES FILHO, Naercio Aquino; PINTO, Cristine Campos de Xavier (Ed.). **Avaliação econômica de projetos sociais**. 3. ed. São Paulo: Itaú Social, 2017.

78 MURAD, M. Hassan et al. A proposed new evidence pyramid. **BMJ Evidence-Based Medicine**, v. 21, n. 4, p. 125–127, 2016.

79 NEUMAN, W. L. **Social Research Methods: Qualitative and Quantitative Approaches**. 7. ed. Harlow: Pearson Education Limited, 2014.

80 ONU MULHERES. **Evaluation Handbook: How to manage gender-responsive evaluation**. [S.l.], 2014. Disponível em: <https://lac.unwomen.org>. Acesso em: 8 abr. 2025.

81 ORDÓÑEZ-MATAMOROS, G. et al. La evaluación ejecutiva y la toma de decisiones de política pública: el caso de la E2+SGR en Colombia. **OPERA**, n. 22, p. 53–79, 2018. Disponível em: <https://revistas.uexternado.edu.co/index.php/opera/article/view/5439>. Acesso em: 4 mar. 2024.

82 ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Better criteria for better evaluation: Revised evaluation criteria definitions and principles for use**. Paris, 2019. Disponível em: <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf>. Acesso em: 4 mar. 2024.

83 _____. **Cost-benefit analysis and the environment: further developments and policy use**. Paris: OECD Publishing, 2018. Disponível em: <https://www.oecd.org/environment/tools-evaluation/cost-benefit-analysis.htm>. Acesso em: 11 abr. 2025.

84 _____. **Evaluation and aid effectiveness: Glossary of key terms in evaluation and results based management**. 2. ed. Paris: OECD Publishing, 2023.

85 _____. **Implementation Toolkit for the OECD Recommendation on Public Policy Evaluation**. Paris, 2025. Disponível em: <https://www.oecd.org/en/publications/implementation-toolkit-for-the-oecd-recommendation-on-public-policy-evaluation_77faa4fe-en.html>. Acesso em: 28 abr. 2025.

86 _____. **Recomendação sobre avaliação de políticas públicas**. Paris, 2022.

87 ORGANIZAÇÃO PARA COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Applying evaluation criteria thoughtfully**. Paris, 2021. Disponível em: <https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/03/applying-evaluation-criteria-thoughtfully_45a54ea7/543e84ed-en.pdf>. Acesso em: 8 abr. 2025.

88 _____. **Evaluating development co-operation: summary of key norms and standards**. 2. ed. Paris: OECD Publishing, 2010.

89 ORGANIZAÇÃO PARA COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Glossary of key terms in evaluation and results based management**. Paris, 2002. Disponível em: <https://www.oecd.org/content/dam/oecd/en/publications/reports/2002/08/evaluation-and-aid-effectiveness-no-6-glossary-of-key-terms-in-evaluation-and-results-based-management-in-english-french-and-spanish_g1gh2dcb/9789264034921-en-fr.pdf#page40>. Acesso em: 8 abr. 2025.

90 _____. **Improving Governance with Policy Evaluation: Lessons From Country Experiences**. Paris: OCDE, 2020. Disponível em: <https://www.oecd.org/en/publications/improving-governance-with-policy-evaluation_89b1577d-en.html>. Acesso em: 9 abr. 2025.

91 PARKHURST, Justin. **The politics of evidence: from evidence-based policy to the good governance of evidence**. Abingdon: Routledge, 2017.

92 PATTON, Michael Quinn. **Developmental evaluation: applying complexity concepts to enhance innovation and use**. New York: Guilford Press, 2011.

93 _____. **Qualitative research & evaluation methods: Integrating theory and practice**. 4. ed. Thousand Oaks: Sage Publications, 2015.

94 _____. **Utilization-focused evaluation**. 4. ed. Thousand Oaks: SAGE Publications, 2008.

95 PAWSON, Ray; TILLEY, Nicholas. **Realistic Evaluation**. Londres: Sage, 1997.

96 PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. **Handbook on planning, monitoring and evaluating for development results**. New York, 2009. Disponível em: <https://www.undp.org/publications/handbook-planning-monitoring-and-evaluating-development-results>. Acesso em: 11 abr. 2025.

97 PROGRAMA MUNDIAL DE ALIMENTOS (PMA). **State of School Feeding Worldwide 2020**. Roma, 2020. Disponível em: <https://alimentacaoescolar.org.br/wp-content/uploads/2024/08/WFP_State_of_School_Feeding_Worldwide__20.pdf>. Acesso em: 11 abr. 2025.

98 REINO UNIDO. TESOURO NACIONAL BRITÂNICO. **The Green Book: appraisal and evaluation in central government**. [S.l.], 2022. Disponível em: <https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent>. Acesso em: 3 abr. 2025.

99 _____. **The Magenta Book: Guidance for evaluation**. Londres, 2020. Disponível em: <https://www.gov.uk/government/publications/the-magenta-book>. Acesso em: 11 abr. 2025.

100 ROBERTS, D.; KHATTRI, N. **Designing a results framework for achieving results: a how-to guide**. Washington, DC, 2012. Disponível em: <https://documents1.worldbank.org/curated/en/331541563854787772/pdf/Designing-a-Results-Framework-for-Achieving-Results-A-How-to-Guide.pdf>. Acesso em: 4 abr. 2025.

101 RODRIGUEZ-BILELLA, P.; TAPELLA, E. (Ed.). **Evaluación, democracia y transformación: Experiencias de evaluación participativa en América Latina**. San Juan: Vientosur, 2024.

102 ROGERS, P. Using programme theory to evaluate complicated and complex aspects of interventions. **Evaluation**, v. 14, n. 1, p. 29–48, 2008. Disponível em: <https://journals.sagepub.com/doi/epdf/10.1177/1356389007084674>. Acesso em: 8 abr. 2025.

103 ROGERS, P.; FUNNELL, S. **Purposeful program theory: effective use of theories of change and logic models**. San Francisco: Jossey-Bass, 2011.

104 ROGERS, Patricia; WOOLCOCK, Michael. Evaluation for adaptive governance. **Journal of Development Effectiveness**, v. 15, n. 2, p. 117–134, 2023.

105 ROSSI, Peter H.; LIPSEY, Mark W.; FREEMAN, Howard E. **Evaluation: a systematic approach**. 7. ed. Thousand Oaks: Sage Publications, 2004.

106 SCARMO, Stephanie; MEAGHER, Whitney. 6 Takeaways Show Strong Progress on School Food and Nutrition. The Pew Charitable Trusts, 2017. Disponível em: <https://www.pewtrusts.org/en/research-and-analysis/articles/2017/04/11/6-takeaways-show-strong-progress-on-school-food-and-nutrition>. Acesso em: 29 abr. 2025.

107 SECCHI, L. **Políticas públicas: conceitos, esquemas de análise, casos práticos**. São Paulo: Cengage Learning, 2014.

108 SNILSTVEIT, Birte et al. **Interventions for improving learning outcomes and access to education in low- and middle-income countries: a systematic review**. London, 2015. Disponível em: <https://www.3ieimpact.org/sites/default/files/2019-05/SR24-education-review_2.pdf>. Acesso em: 3 abr. 2025.

109 STETSON, Valerie. **Communicating and reporting on an evaluation: Guidelines and tool**. Washington, DC, 2008. Disponível em: <https://www.betterevaluation.org/sites/default/files/MEmodule_communicating.pdf>. Acesso em: 30 abr. 2025.

110 VAESSEN, J. **Mind the Gap: Perspectives on Policy Evaluation and the Social Sciences**. Nova York: Routledge, 2017. Disponível em: <https://www.taylorfrancis.com/books/edit/10.4324/9781315124537/mind-gap-jos-vaessen?refId=7ae7be14-6eab-4944-94d2-b6ec4e58a268&context=ubx>. Acesso em: 11 abr. 2025.

111 VAESSEN, Joss; LEMIRE, Sebastian; BEFANI, Barbara. **Evaluation of International Development Interventions: An Overview of Approaches and Methods**. Washington, DC, 2020. Disponível em: <https://documents1.worldbank.org/curated/en/942511608652015232/pdf/Evaluation-of-International-Development-Interventions-An-Overview-of-Approaches-and-Methods.pdf>. Acesso em: 14 abr. 2025.

112 VENEKLASEN, L.; MILLER, V. Chapter 9: Planning Moment #4: Analyzing Problems & Selecting Priority Issues. In: A new weave of power, people & politics: the action guide for advocacy and citizen participation. [S.l.]: Just Associates, 2007. Disponível em: <https://justassociates.org/all-resources/a-new-weave-of-power-people-politics-the-action-guide-for-advocacy-and-citizen-participation/>. Acesso em: 31 ago. 2023.

113 WASHINGTON STATE INSTITUTE FOR PUBLIC POLICY (WSIPP). **Benefit-Cost Technical Documentation: Washington State Institute for Public Policy Benefit-Cost Model**. Olympia, 2024. Disponível em: <https://www.wsipp.wa.gov/TechnicalDocumentation/WsippBenefitCostTechnicalDocumentation.pdf>. Acesso em: 17 abr. 2025.

114 WATKINS, R.; WEST MEIERS, M.; VISSER, Y. L. **A guide to assessing needs: essential tools for collecting information, making decisions, and achieving development results**. Washington, DC: The World Bank, 2012. Disponível em: <http://hdl.handle.net/10986/2231>. Acesso em: 31 ago. 2023.

115 WEISS, C. Theory-based evaluation: past, present, and future. **New Directions for Evaluation**, n. 76, p. 41–55, 1997. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/ev.1086>. Acesso em: 8 abr. 2025.

116 WHITE, Howard. A contribution to current debates in impact evaluation. **Evaluation**, v. 16, n. 2, p. 153–164, 2010.

117 WHOLEY, Joseph S.; HATRY, Harry P.; NEWCOMER, Kathryn E. (Ed.). **Handbook of practical program evaluation**. 3. ed. São Francisco: Jossey-Bass, 2010.

118 WILLIAMS, Eleanor. **Rapid Evaluation**. [S.l.], 2022. Disponível em: <https://www.betterevaluation.org/methods-approaches/approaches/rapid-evaluation>. Acesso em: 15 abr. 2025.

119 ZHUANG, Juzhong et al. **Theory and practice in the choice of social discount rate for cost-benefit analysis: a survey**. Manila: Asian Development Bank, 2007.

Apêndice I - Ferramentas baseadas em inteligência artificial para apoiar a busca e síntese de evidências

Soluções baseadas em inteligência artificial podem ser aliadas poderosas no trabalho avaliativo, especialmente quando há necessidade de localizar e organizar evidências em curto prazo ou em contextos com ampla produção científica. Embora seu uso ainda seja recente na avaliação de políticas públicas, essas soluções já vêm sendo exploradas com bons resultados para apoiar a revisão de evidências e a sistematização de achados relevantes.

Tabela 12: Ferramentas baseadas em IA para revisão de literatura científica

Ferramenta	Possibilidades de uso	Link
ChatGPT + Plugins (ScholarAI / Scispace)	Permite fazer perguntas abertas e obter respostas baseadas em artigos científicos. Com plugins ativados, busca DOIs, resume papers e ajuda a escrever revisões com base nas evidências encontradas.	https://chat.openai.com
Scite.ai	Mostra como outros estudos citam um artigo (se concordam, discordam ou apenas mencionam), facilitando a filtragem de estudos robustos.	https://scite.ai
Elicit (by Ought)	Ideal para revisões rápidas. Exibe tabela com artigos relevantes, autores, métodos (RCT, quase-experimentos), países e achados principais.	https://elicit.org
Research Rabbit	Mapeia redes de autores e temas. Útil para identificar grupos de pesquisa e conexões entre estudos.	https://www.researchrabbit.ai
Connected Papers	Gera um mapa visual de artigos relacionados a um paper inicial. Útil para construir revisões conectadas.	https://www.connectedpapers.com
Consensus	Responde perguntas abertas com base em artigos científicos revisados por pares.	https://consensus.app

Fonte: FGV CLEAR, elaboração própria.

Como usar essas ferramentas para revisar evidências em políticas públicas?

- Comece pelo Elicit (https://elicit.org)
  → Digite sua pergunta. Veja uma tabela com artigos relevantes, autores,

países estudados, métodos (RCT, quase-experimental etc.) e achados principais.

- Explore os autores e papers mais citados no Research Rabbit (https://www.researchrabbit.ai)

  Aprofunde sua busca por rede: quem são os especialistas no tema? Há grupos de pesquisa latino-americanos?

- Use o Scite.ai para validar a força dos estudos encontrados (https://scite.ai)

  → Veja se o estudo foi citado por outros como evidência forte, questionado, ou pouco referenciado.

- Use o ChatGPT com plugin ScholarAI ou Scispace para sintetizar e redigir (https://chat.openai.com)

  → Exemplo de comando: "Com base nesses artigos, escreva uma revisão da literatura sobre políticas eficazes de inserção produtiva para mulheres na América Latina."

- Verifique achados convergentes com o Consensus (https://consensus.app)
  → Teste perguntas mais fechadas para checar consenso acadêmico (ex: "Microcrédito melhora renda de mulheres?").

Apesar de serem aliadas valiosas na busca e organização de evidências, as ferramentas de inteligência artificial têm limitações que merecem atenção. Nem sempre conseguem acessar todos os estudos relevantes, especialmente quando os artigos estão em bases fechadas ou fora dos repositórios mais conhecidos. Além disso, os resumos e respostas gerados podem omitir aspectos importantes do contexto ou da metodologia dos estudos, o que pode levar a interpretações apressadas. Por isso, o ideal é utilizá-las como ponto de partida — elas ajudam a ganhar tempo e a organizar o que já existe, mas não substituem a leitura crítica dos textos originais nem a escuta de especialistas da área.