

Guia de Avaliação de Impacto Socioambiental

para Utilização em Projetos e Investimentos de Impacto



Insper METRICiS

Núcleo de Medição para Investimentos de Impacto Socioambiental

Sobre o Insper Metricis

O Insper Metricis é um centro de pesquisa focado em realizar estudos sobre estratégias organizacionais e práticas de gestão envolvendo projetos com o potencial de gerar alto impacto socioambiental. Ênfase especial é dada ao desenvolvimento de ferramentas para planejar, executar e avaliar projetos de impacto realizados por empresas, organizações sem fins lucrativos e governos. Com procedimentos de gestão e avaliação sobre a potencial contribuição de projetos de cunho socioambiental, é possível estimular não somente o crescimento dos investimentos em projetos de impacto, como também impulsionar o desenvolvimento de novas maneiras de captação e financiamento. Além disso, o aprendizado gerado por essas experiências permite o contínuo registro e a disseminação de melhores práticas, por meio de pesquisas acadêmicas, relatórios de políticas públicas, estudos de casos e guias de gestão ligados a projetos de alto impacto.

Sobre o Insper

O Insper é uma instituição independente e sem fins lucrativos que se dedica ao ensino e à pesquisa nas áreas de Administração, Economia, Direito, Ciência da Computação e Engenharia. Tem como missão ser um centro de referência e explorar as complementariedades dessas áreas. Suas atividades de ensino abrangem cursos para várias etapas de uma trajetória profissional: graduação (Administração, Economia, Direito, Ciência da Computação e Engenharias), pós-graduação lato e stricto sensu (Certificates, MBAs, programas na área de Direito, mestrados profissionais e doutorados) e educação executiva (programas abertos e customizados de acordo com as necessidades das empresas). No âmbito da produção de conhecimento, o Insper atua por meio de cátedras e centros de pesquisa que reúnem pesquisadores em estudos e projetos focados em políticas públicas, finanças e gestão. A instituição conta ainda com centros que promovem o empreendedorismo (CEMP) e educação (Centro de Educação). Tem as certificações de qualidade da Association to Advance Collegiate Schools of Business (AACSB), Association of MBAs (AMBA) e Associação Nacional de MBAs (Anamba). O Insper Metricis é um centro de pesquisa ligado ao Centro de Gestão e Políticas Públicas (CGPP).

São Paulo

Quinta edição, maio de 2022

Guia geral com foco em monitoramento e verificação de adicionalidade

A versão original deste documento foi escrita por Sérgio G. Lazzarini, Leandro S. Pongeluppe, Pui Shenn Yoong e Nobuiuki Costa Ito. Nesta nova edição, a revisão foi elaborada por Sérgio G. Lazzarini, José Geraldo Setter Filho, Carolina Pedrosa Gomes de Melo, Jorge Norio Rezende Ikawa, Octavio Augusto Darcie de Barros e Carolina Pellegrino Castejon, e recebeu sugestões de Lígia Vasconcellos. Edições anteriores contaram com as contribuições de Amanda Arabage, Sandro Cabral, Sergio Firpo, Pedro Machado de Godoy, Guilherme Lichand, Carolina Pedrosa Gomes de Melo, Marina Ribeiro, José Geraldo Setter Filho, Mariana Suplicy, Carlos Kazunari Takahashi e Rafael Vivolo, além de sugestões dadas por Fernando Carnaúba, Ben Carpenter, Amanda Feldman, Naercio Menezes Filho, Ricardo Paes de Barros, Luís Fernando Guedes Pinto, Brian Trelstad e Mauricio Voivodic. Discussões iniciais para desenvolver este guia evoluíram a partir do trabalho de um grupo envolvendo Raquel Costa, Célia Cruz, Tatiana Fonseca, Frédéric Kuonen, Simon Locher, Angélica Rotondaro, Franco Veludo e Rafaella Ziegert, contando com o apoio do Latin America Economy Impact Innovations Fund (The Rockefeller Foundation, Fundación Avina e Omidyar Network), em proposta coordenada pelo Instituto de Cidadania Empresarial (ICE), por meio de Célia Cruz e Maria Amélia Sampaio.

Introdução ao guia



Muitas organizações privadas, públicas e sem fins lucrativos buscam avaliar se seus projetos são geradores de melhorias consistentes e mensuráveis na vida de sua população-alvo. Em muitos casos, patrocinadores e beneficiários desses projetos também gostariam de saber se as intervenções propostas geram benefícios, para que, dessa forma, saibam quando apoiar iniciativas bem-sucedidas e quando descontinuar projetos com falta de evidências de impacto positivo.

Há também uma preocupação crescente de que organizações que busquem demonstrar rapidamente seu alinhamento com práticas socioambientais (seguindo a nova tendência de “ESG”) possam relatar incorretamente seu real impacto e apresentar análises superficiais apenas para anunciar benefícios sociais que, muitas vezes, não podem ser verificados (um problema conhecido como “*greenwashing*” ou “*impact washing*”).

Até mesmo em organizações que selecionam e seguem métricas apropriadas de desempenho socioambiental, geralmente, há uma confusão a respeito do impacto efetivamente causado na população-alvo e se as melhorias reivindicadas foram realmente efetivas. Nesse âmbito, também é importante distinguir o **monitoramento** do projeto de sua **avaliação de impacto**.

Usualmente, os gestores de projetos coletam indicadores da população-alvo e monitoram esses dados ao longo do tempo. Por exemplo, em um projeto pensado para aumentar o bem-estar de comunidades rurais carentes, os gerentes do projeto podem seguir indicadores de renda, educação, saúde e vários outros. A partir dessas informações, metas podem ser estabelecidas, e medidas corretivas podem ser introduzidas quando os resultados reais estiverem abaixo do nível esperado.

No entanto, para efetivamente avaliar se o projeto *causou* mudanças nessas métricas, os gerentes precisariam analisar o impacto do projeto, respondendo a seguinte pergunta: “o que teria acontecido com as comunidades ou os indivíduos-alvo caso eles *não* tivessem sido expostos ao projeto?”

A resposta a essa pergunta é importante porque melhorias externas podem ter acontecido simultaneamente ao projeto, de modo a afetar a população-alvo de forma independente. Por exemplo, em uma iniciativa de fornecimento de ferramentas tecnológicas de aprendizagem para escolas públicas, o gerente do projeto pode concluir incorretamente que as ferramentas melhoraram a aprendizagem do aluno quando, na verdade, a melhoria pode ter sido causada por mudanças nas práticas pedagógicas da rede pública de ensino como um todo.

A chamada verificação de **adicionalidade** em projetos de cunho social tenta evitar justamente esse tipo de avaliação equivocada. Nessa abordagem, o impacto é definido pela diferença entre o resultado observado no indivíduo que participou do projeto e o resultado que teria sido observado nesse mesmo indivíduo, durante o mesmo período de tempo, caso a intervenção não tivesse sido implementada — chamado de **contrafactual**.

Sendo assim, idealmente precisaríamos comparar os dois estados da *mesma* pessoa – ou do mesmo grupo – que recebe a intervenção. Em alguns casos, isso é viável. Por exemplo, é possível medir instantaneamente o efeito do uso de óculos na capacidade de leitura de pessoas com deficiência visual, bastando uma simples comparação entre quanto a pessoa pode ver com e sem óculos.

Na maior parte dos casos, contudo, os resultados das intervenções ocorrem por um longo período de tempo e podem ser afetados por vários fatores que não estão diretamente relacionados ao projeto. Nesses casos, a pergunta “o que teria acontecido com as comunidades ou indivíduos-alvo caso eles *não* tivessem sido expostos ao projeto?” pode ser respondida com base em informações de pessoas que, simultaneamente aos beneficiários dos projetos, *não* receberam a intervenção. No entanto, deve-se considerar uma estratégia de avaliação que permita controlar os fatores diversos que podem afetar tanto o tratamento como os resultados.

Por exemplo, em estudos de eficácia médica, o **grupo de tratamento** (geralmente escolhido de modo aleatório) recebe um medicamento ou um tratamento médico, enquanto o **grupo de controle** não os recebe. Isso permite medir como a intervenção médica afetou a população, levando-se em consideração as mudanças naturais que poderiam ter ocorrido. Nesse caso, por conta da atribuição aleatória, o grupo de controle funciona como o contrafactual. Quando um experimento aleatório não pode ser implementado, outras estratégias de medição, baseadas na adicionalidade, podem ser usadas para simular o contrafactual da melhor forma possível.

A **Figura 1** exemplifica esse tipo de abordagem de medição baseada na adicionalidade. Imagine que uma empresa queira investir no desenvolvimento de diversas comunidades carentes e avaliar o impacto dos investimentos correspondentes. Assim, a empresa define as comunidades que serão afetadas pelo projeto (o grupo de tratamento) e as comunidades que não receberão as intervenções (o grupo de controle). Antes de iniciar o projeto, esses dois grupos apresentam faixas de renda diferentes, mas, com relação à métrica de resultados estabelecida, têm padrões de crescimento similares.

No início do projeto, a empresa mede o nível de renda dos indivíduos dos grupos de controle e tratamento. Nesse momento, a renda média do grupo de controle é de \$90, enquanto a renda média do grupo de tratamento (que receberá o investimento) é de \$100.

Após fazer essas medições, a empresa inicia o projeto com a comunidade. Depois de um ano, ela mede novamente a renda média dos grupos de controle e de tratamento. Um analista inexperiente pode concluir que o impacto do projeto é de \$30, já que a renda média do grupo de tratamento aumentou de \$100 para \$130. No entanto, utilizando a verificação de adicionalidade, a evolução do grupo de tratamento é comparada ao que provavelmente teria acontecido sem o projeto. O cenário contrafactual, nesse caso, é representado pelo que aconteceu com o grupo de controle.

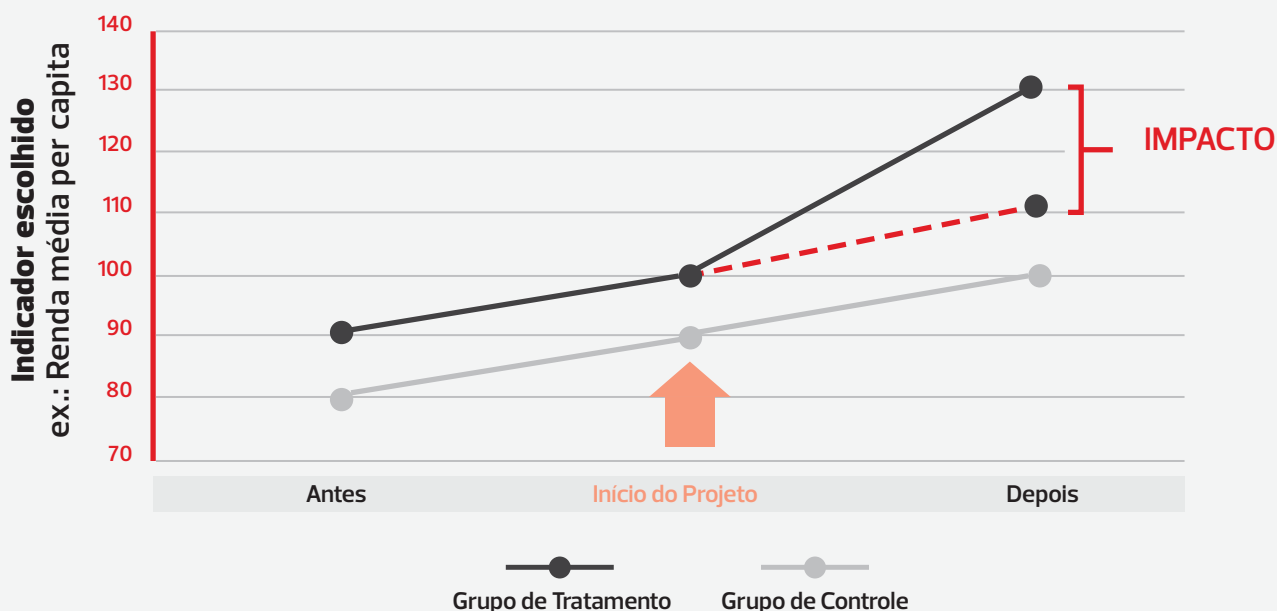
Na maior parte dos casos, os resultados das intervenções ocorrem por um longo período de tempo e podem ser afetados por vários fatores que não estão diretamente relacionados ao projeto.



Devido a outros fatores que podem ter afetado a comunidade (como uma evolução natural nos padrões de vida ou investimentos do governo), o grupo que não recebeu o investimento do projeto também teve um aumento de renda de \$10 por indivíduo. Medir o impacto real do projeto é possível se não limitarmos a análise às mudanças ao longo do tempo apenas no grupo de tratamento, levando em conta também outras tendências naturais que podem ocorrer (conforme evidenciado pelo grupo de controle).

Nesse caso, está claro que o projeto aumentou em \$20 a renda per capita média dos indivíduos da comunidade de tratamento. Esse ganho adicional é simplesmente a diferença entre a mudança verificada na comunidade de tratamento (\$30) e a mudança verificada na comunidade de controle (\$10). O aumento de renda neste último grupo serve como uma estimativa do cenário contrafactual (o que teria acontecido à população-alvo sem o projeto).

Figura 1 – Avaliação de impacto por meio da verificação de adicionalidade



Nesse sentido, o presente *Guia* fornece um conjunto estruturado de etapas para desenvolver planos de medição de impacto, com dois objetivos principais:

- 1. Propor métricas altamente relevantes para avaliar potenciais melhorias que sejam valorizadas pela população-alvo do projeto.**
- 2. Avaliar o que teria acontecido com a população-alvo na ausência da intervenção (geralmente por meio de comparações com grupos de controle similares, mas que não estejam incluídos no projeto).**

Enquanto o objetivo 1 geralmente é utilizado para fins de monitoramento, o objetivo 2 vai um passo além e busca avaliar o impacto causado pelo projeto. Nesse contexto, o *Guia de Avaliação de Impacto Socioambiental para Utilização em Projetos e Investimentos de Impacto* fornece uma ferramenta prática para permitir que investidores de impacto, organizações sem fins lucrativos e governos monitorem e avaliem os resultados potencialmente gerados por meio de seus projetos.

O Guia propõe um passo-a-passo, partindo da definição dos objetivos do projeto e sua população-alvo, acompanhado de uma sequência de ferramentas para selecionar as métricas adequadas e encontrar formas alternativas de avaliar o impacto causado pelo projeto (ou seja, sua adicionalidade).

Nesse sentido, o *Guia* é uma ferramenta útil para a prática da gestão de impacto. Por exemplo, seguindo a abordagem proposta pelo *Impact Management Project (IMP)*¹, o *Guia* fornece uma discussão focada na dimensão de “**contribuição**”, isto é, até que ponto os resultados foram distintos daqueles que teriam acontecido sem o projeto. No entanto, alinhado com a estrutura geral do IMP, o *Guia* também discute **quem** são os potenciais beneficiários (definição da população-alvo), **quais** são os resultados mais relevantes (definição da teoria da mudança e seleção de métricas) e **quanto** o projeto pode melhorar o bem-estar de sua população-alvo (baseado em resultados de estudos anteriores e em resultados esperados de acordo com as intervenções planejadas).



O Guia é uma ferramenta útil para a prática da gestão de impacto.

¹ Ver: IMPACT MANAGEMENT PROJECT. *Mainstreaming the practice of impact management*. Disponível em: <https://impactmanagementproject.com/>. Acesso em: 9 mar. 2022.

Preparando um plano de medição



DESCRIÇÃO DO PROJETO
E DE SEUS OBJETIVOS
MAIS AMPLOS



DEFINIÇÃO DA
POPULAÇÃO-ALVO



BENCHMARKING



DEFINIÇÃO DA
TEORIA DA MUDANÇA



DEFINIÇÃO
DE MÉTRICAS



DEFINIÇÃO DO MÉTODO
PARA VERIFICAR
ADICIONALIDADE



PLANO DE
AMOSTRAGEM



CRONOGRAMA
DE MEDIÇÃO

Preparando um plano de medição

Preparar um plano de ação é uma etapa essencial para configurar, operacionalizar, implementar e controlar adequadamente a medição. Para implementar medições de dados robustas e precisas, as etapas a seguir devem ser consideradas. As etapas de 1 a 5 permitem a criação de *scorecards* de monitoramento compostos de métricas adequadas para medir desempenho socioambiental. As etapas de 6 a 8, por sua vez, envolvem a avaliação do impacto do projeto (ou seja, sua adicionalidade).



1. DESCRIÇÃO DO PROJETO E DE SEUS OBJETIVOS MAIS AMPLOS

A estratégia para definir como estimar o impacto de uma intervenção começa com uma descrição preliminar da organização e das intervenções planejadas. Nesta etapa, o propósito da organização e os objetivos gerais dos projetos propostos devem ser descritos e discutidos.

Na gestão de impacto, tem sido costume vincular os objetivos gerais do projeto aos **Objetivos de Desenvolvimento Sustentável (ODSs)** das Nações Unidas, cada vez mais utilizados para identificar questões fundamentais a ser abordadas como forma de melhorar o bem-estar das populações em todo o mundo². É possível que o mesmo projeto esteja relacionado a vários ODSs. Por exemplo, considere um projeto que tenha como objetivo aumentar o acesso de comunidades carentes à infraestrutura de água potável e saneamento. Esse projeto está muito ligado ao ODS 6: "Garantir a disponibilidade e a gestão sustentável da água potável e do saneamento para todos". No entanto, ao mesmo tempo, o projeto pode impactar positivamente as condições de saúde da população-alvo e promover uma infraestrutura local sustentável, criando, assim, ligações adicionais com os ODSs 3 e 11, respectivamente.

Muitas vezes, mesmo quando as intervenções propostas são muito bem definidas por seus gestores, os objetivos são pouco discutidos e detalhados. Além disso, é comum encontrarmos projetos que envolvam uma miríade de objetivos e intervenções cuja execução requer capacidades que a organização não tem. Portanto, aprofundar a discussão dos objetivos do projeto também pode ajudar os gerentes a entender quais desses objetivos são diretamente influenciados pelo projeto e quais, embora desejáveis e relacionados às ações implementadas, estão fora de seu escopo.



2. DEFINIÇÃO DA POPULAÇÃO-ALVO

Uma etapa essencial na definição de um plano de medição de impacto é determinar a **população-alvo** do projeto. Para a delimitação objetiva desse público, é necessário, em primeiro lugar, identificar todos os atores relevantes que possam ser afetados pelo projeto ou influenciar seus resultados (os chamados **stakeholders**). Esses atores podem incluir financiadores, provedores de serviço, organizações de suporte e, especialmente, os beneficiários das intervenções. É importante, nesta etapa, verificar se suas prioridades e seus objetivos estão de acordo com as ações planejadas para o projeto.

² Ver: NAÇÕES UNIDAS BRASIL. **Objetivos de Desenvolvimento Sustentável**. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 20 mar. 2022.

Como as organizações geralmente têm recursos limitados, a escolha de uma população-alvo também é uma oportunidade de focar quem realmente precisa de ajuda ou de enfatizar os subgrupos para os quais a intervenção pode ser mais efetiva.

Por exemplo, um projeto pode buscar melhorar as condições de saúde de populações vulneráveis, aumentando seu acesso a medicamentos e métodos eficazes para o tratamento de doenças crônicas a um custo mais baixo. No entanto, quais são essas populações carentes, e quais indivíduos poderão ter prioridade, especialmente em uma condição de escassez de recursos?

Fatores possíveis de ser considerados no processo de definição da população-alvo incluem :³

- **Geografia:** a delimitação regional do projeto;
- **Características demográficas:** idade, gênero, raça e outras;
- **Características socioeconômicas:** renda, escolaridade, ocupação e outras;
- **Necessidades especiais:** deficiências físicas, cognitivas e outras vulnerabilidades relevantes;
- **Susceptibilidade a fatores ambientais:** clima adverso, desastres naturais e outros;
- **Condições iniciais:** como o segmento focalizado encontra-se em termos de variáveis de interesse do projeto.

O **Box 1** apresenta um exemplo de focalização da população-alvo considerando alguns dos fatores acima indicados.

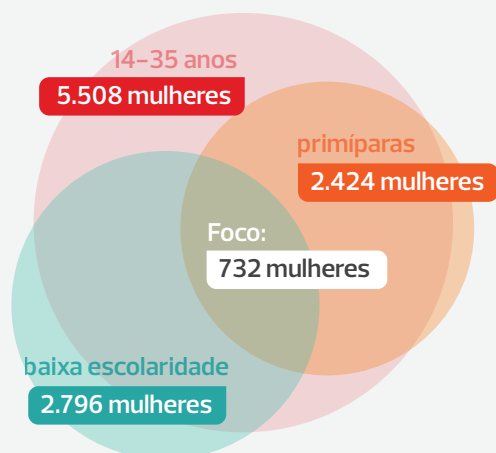
BOX 1. Exemplo de focalização da população-alvo em um contexto de ações pré-natal com mulheres grávidas.

Um programa governamental em parceria com uma organização sem fins lucrativos busca realizar consultas de saúde pré-natal para melhorar os indicadores de saúde materna e dos recém-nascidos. O governo definiu um conjunto de municípios, totalizando 6.024 mulheres grávidas.

Estudos anteriores indicam que a eficácia desses programas é maior no caso de mulheres vulneráveis com idade entre 14 e 35 anos (*característica demográfica*), com até sete anos de escolaridade (*característica socioeconômica*) e em primeira gravidez (*condição inicial*).

O diagrama de Venn ao lado indica a quantidade de mulheres com cada uma dessas características e o total de casos na sua intersecção: 732 mulheres. Essas mulheres poderiam ser consideradas como prioritárias para as intervenções propostas, ainda que outros segmentos também possam ser incluídos no programa, caso existam recursos suficientes.

Total de grávidas nos municípios do programa: 6.024



³ Para mais detalhes, ver: IMPACT MANAGEMENT PROJECT. **Who**. Disponível em: <https://impactmanagementproject.com/impact-management/impact-management-norms/who/>. Acesso em: 3 jun. 2019.



3. BENCHMARKING



A expansão dos projetos de impacto depende de um esforço de aprendizado contínuo baseado nos erros e acertos de projetos similares, bem como no aproveitamento de medições obtidas por pesquisas previamente realizadas. Portanto, o *benchmarking* é uma etapa essencial para identificar o que já foi feito ou estudado em relação à atividade e à população-alvo do projeto.

Fontes úteis de informação incluem estudos acadêmicos, publicações de organizações internacionais, relatórios de outros patrocinadores de projetos na mesma atividade, entre outras. Há ainda diversas plataformas na internet que consolidam estudos de investigação e avaliação em várias áreas, com informações detalhadas sobre possíveis intervenções e sua efetividade⁴. Em estudos de impacto, também é cada vez mais comum consultar *meta-análises*: estudos que buscam compilar os resultados de várias avaliações anteriores sobre um determinado tema.

Como exemplo, imagine um gestor interessado em ajudar pessoas desempregadas a encontrar boas oportunidades de trabalho (assim, conectando-se ao ODS 8). Uma meta-análise publicada em 2008 analisou 207 estudos sobre programas ativos de mercado de trabalho, incluindo treinamento e intermediação de empregos, bem como subsídios para empresas⁵. O estudo mostra que os resultados observados após um ano de intervenção são superiores aos de programas voltados para o curto prazo (até um ano). Comparando-se os tipos de intervenção, os programas com ênfase em treinamento têm resultados mais efetivos do que outros programas, que fazem apenas intermediações com as empresas.

A etapa de *benchmarking* também permite quantificar o tamanho do efeito esperado para o projeto. Por exemplo, na meta-análise citada anteriormente, os autores concluem que os programas de treinamento tendem a aumentar a probabilidade de emprego em até 8,7 pontos percentuais no longo prazo. Esse efeito é superior ao observado após apenas um ano (1,6 ponto percentual), o que, por si só, é uma indicação de que é desejável avaliar o impacto do programa durante um período de tempo mais longo. Caso a análise dos programas anteriores indique efeitos pouco relevantes, os gestores também podem tentar mudar ou complementar as intervenções propostas.

É possível ainda que o efeito do projeto seja maior ou menor dependendo das características heterogêneas da população, como discutido na Etapa 2. Ainda usando o exemplo de intervenções no mercado de trabalho, os gestores podem tentar focar jovens de baixa renda que recentemente se formaram no Ensino Médio. O *benchmarking* deveria, então, localizar estudos especificamente focados nessa população-alvo.



⁴ Por exemplo, a biblioteca de avaliações de impacto do Abdul Latif Jameel Poverty Action Lab – J-PAL (J-PAL. **Abdul Latif Jameel Poverty Action Lab**. Disponível em: <https://www.povertyactionlab.org/>. Acesso em: 20 mar. 2022); a ferramenta de busca de intervenções em educação da Education Endowment Foundation (EDUCATION ENDOWMENT FOUNDATION. **Teaching and Learning Toolkit**. Disponível em: <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit>. Acesso em: 20 mar. 2022); e os estudos compilados pela Cochrane Library em saúde (COCHRANE. **Our Evidence**. Disponível em: <https://www.cochrane.org/EVIDENCE>. Acesso em: 20 mar. 2022).

⁵ Ver: CARD, D.; KLUVE, J.; WEBER, A. What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. **Journal of the European Economic Association**, v. 16, n. 3, p. 894–931, 1 jun. 2018.



4. DEFINIÇÃO DA TEORIA DA MUDANÇA



Antes de aprofundar a questão de como medir o impacto do projeto, é necessário definir o que medir. Nesta etapa, os gestores dos projetos devem desenvolver a teoria da mudança: uma forma clara e lógica de conectar as intervenções propostas aos resultados socioambientais esperados.

A **Figura 2** mostra os cinco estágios de uma teoria da mudança, cujo mapeamento é essencial para qualquer plano de medição: **insumos** e **atividades**, que geram **produtos** ofertados à população-alvo; os quais, por sua vez, desencadeiam **resultados** relevantes ligados tanto às atividades principais do projeto, quanto a transformações sociais mais amplas. Esta etapa se beneficia diretamente das etapas anteriores do plano de medição. Por exemplo, *benchmarking* permite identificar quais potenciais atividades e insumos podem ser mais efetivos para gerar esses resultados.

Figura 2 – Teoria da mudança aplicada a investimentos e negócios de impacto



Fonte: Adaptado pelos autores com base em McLaughlin & Jordan (2004)⁶

Geralmente, gestores de projetos socioambientais tendem a reportar produtos em vez de resultados. Para ilustrar essa distinção, considere uma empresa que tenha como principal atividade disponibilizar ferramentas tecnológicas para melhorar o desempenho escolar dos alunos da rede pública de ensino, utilizando diversos insumos, como plataformas on-line e equipes de apoio para coordenar as intervenções. Os produtos dessa atividade seriam o número de acessos às ferramentas ou o número de sessões interativas com os alunos. Observe, no entanto, que esses produtos não indicam necessariamente que os alunos melhoraram como resultado da intervenção. É preciso, portanto, analisar os resultados, ou seja, as mudanças relevantes que causaram impacto positivo na vida dos beneficiários ou na sociedade em geral.

⁶ Ver: MCLAUGHLIN, J. A.; JORDAN, G. B. Using logic models. In: NEWCOMER, K. E.; HATRY, H. P.; WHOLEY, J. S. (Eds.). **Handbook of practical program evaluation**. 4. ed. New Jersey: Jossey-Bass, 2015. p. 62–87. Ver também: HEHENBERGER, L.; HARLING, A.-M.; SCHOLTEN, P. **A Practical Guide to Measuring and Managing Impact**. Brussels: European Venture Philanthropy Association, 2013. Disponível em: https://www.oltventure.com/wp-content/uploads/2015/05/EVPA_A_Practical_Guide_to_Measuring_and_Managing_Impact_final.pdf. Acesso em: 20 mar. 2022.

Nesse caso, um resultado esperado da atividade é a melhoria nas notas dos alunos em um determinado teste padronizado. O resultado social do projeto, por sua vez, seria o bem-estar futuro e até mesmo o ganho de renda futuro que os alunos poderão colher como resultado de suas habilidades adicionais. Nesse processo, é importante descrever e justificar o **mecanismo causal** que influenciará a mudança desejada. Por exemplo, as ferramentas tecnológicas podem incluir jogos e vídeos, potencialmente aumentando a motivação dos alunos para aprender e promovendo sua capacidade de compreender e absorver novos conhecimentos.

A teoria da mudança também permite a criação de um *scorecard* de métricas relevantes a ser monitoradas, com base nos dados dos beneficiários do projeto. No exemplo acima, a empresa pode acompanhar o total de horas despendidas pela equipe de suporte (insumos), as ações que realizam com esses alunos (atividades), o número de alunos que acessaram a ferramenta de aprendizagem (produtos) e a aprendizagem avaliada (resultados), entre outros indicadores.

Em algumas representações da teoria da mudança, a última etapa da análise é chamada de "impacto". No entanto, sob a perspectiva de adicionalidade, a teoria da mudança não é, em si, um método de avaliação de impacto. O impacto do projeto, conforme discutido antes e detalhado na Etapa 5, envolve mudanças que foram causadas exclusivamente pelo projeto, e não por outros fatores externos que afetem a população-alvo.

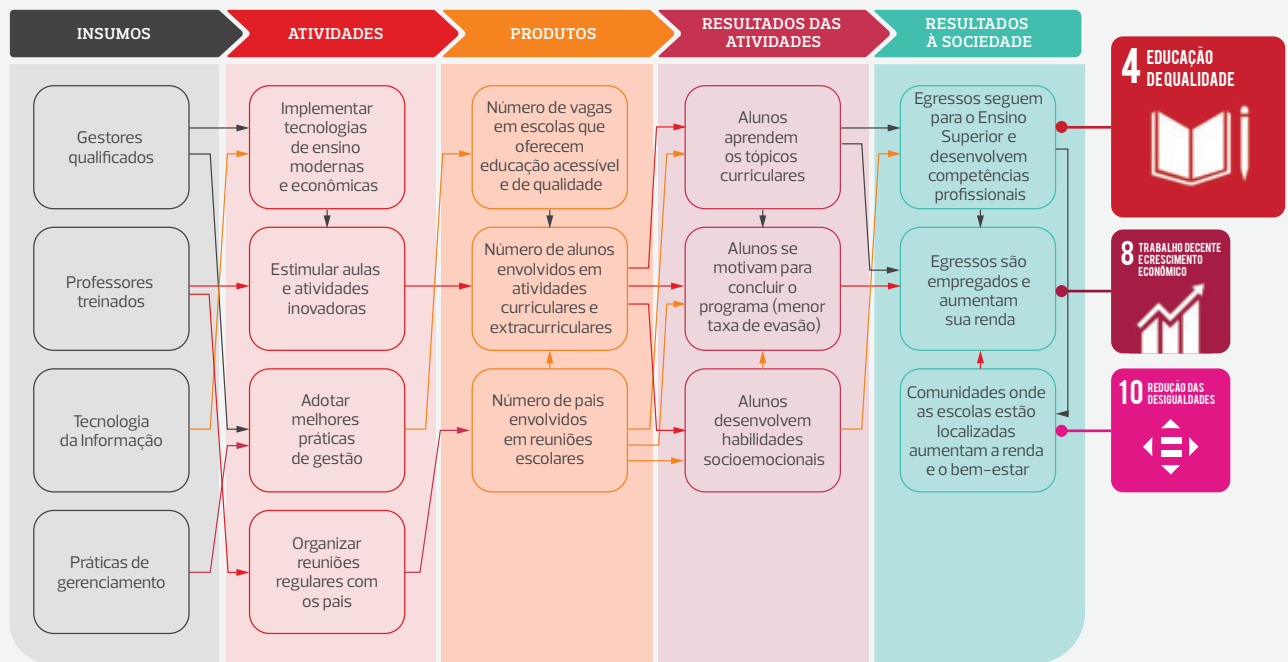
O **Box 2** apresenta outro exemplo de teoria da mudança no campo educacional, que inclui um conjunto mais complexo de insumos e intervenções, além de especificações mais detalhadas das ligações de causa e efeito entre os blocos de construção da teoria. No exemplo abaixo, um investidor de impacto adquiriu um grupo de escolas privadas voltadas para famílias de baixa renda e buscou um conjunto de práticas pedagógicas e gerenciais para melhorar o aprendizado e reduzir a evasão nas escolas secundárias.

Observe, em particular, como os componentes da teoria da mudança são escritos. As atividades identificam as ações que os gerentes realizam (observe os verbos "implementar", "adotar" e outros). As saídas envolvem, em geral, mudanças em números e quantidades. Os resultados, por sua vez, identificam como a população-alvo melhora em função das intervenções propostas ("alunos *aprendem* os tópicos curriculares").



Em algumas representações da teoria da mudança, a última etapa da análise é chamada de "impacto". No entanto, sob a perspectiva de adicionalidade, a teoria da mudança não é, em si, um método de avaliação de impacto.

BOX 2. Exemplo de teoria da mudança para uma rede de escolas secundárias voltada a alunos de baixa renda.



A teoria da mudança pode ser conectada ao Passo 1 do plano de medição, indicando como os resultados à sociedade propostos estão vinculados aos objetivos do projeto e aos ODSs. No **Box 2**, o ícone que simboliza o ODS 4 ("Educação de Qualidade") é maior do que os outros porque representa o objetivo central do projeto, relacionado com a melhoria dos resultados educacionais nas escolas secundárias. No entanto, ao aumentar o acesso dos alunos à educação de alta qualidade, o projeto também os ajuda a encontrar bons empregos no futuro e, potencialmente, reduz as desigualdades de renda nas comunidades-alvo.

Ao definir a teoria da mudança, é importante levar em consideração todos os resultados possíveis, tanto positivos quanto negativos. Negligenciar resultados negativos aumenta o risco de gerar resultados indesejados. Por exemplo, um projeto para aumentar a renda em comunidades rurais pode estimular novas atividades agrícolas que renovem a renda dos agricultores, mas, ao mesmo tempo, pode causar repercussões negativas ao meio ambiente. Certas atividades também podem criar o risco de violação dos direitos humanos básicos, como em casos de discriminação de gênero ou de condições de trabalho precárias. Uma sugestão é considerar as possíveis interações entre os objetivos gerais do projeto (conforme discutido na Etapa 1), a fim de fornecer uma visão mais ampla de todos os resultados possíveis de ser influenciados pela intervenção.





5. DEFINIÇÃO DE MÉTRICAS



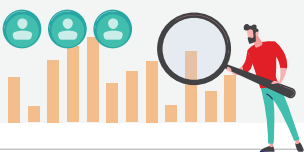
Após concluir os passos anteriores, é possível escolher métricas adequadas para o projeto. Ancorados na teoria da mudança do projeto, os gestores podem desenvolver um painel de métricas para fins de monitoramento. Insumos, atividades e produtos embasam diretamente a criação de um painel *operacional* que acompanha a implementação do projeto. No exemplo do **Box 2**, os gestores podem rastrear a presença de professores qualificados (insumo), a implementação de tecnologias de apoio (atividade) e o número de alunos envolvidos nas atividades da escola (produto).

Embora as métricas possam ser escolhidas para todos os estágios da teoria da mudança, inclusive para fins de monitoramento de atividades e produtos, monitorar e avaliar melhorias fundamentais na população-alvo requerem ênfase nas métricas de resultado. Assim, os resultados podem ser entendidos como *indicadores-chave de desempenho* (KPIs, na sigla em inglês) no painel de monitoramento. O **Box 2** apresenta três KPIs vinculados aos resultados: aprendizagem do aluno de tópicos curriculares, conclusão do programa e desenvolvimento de habilidades socioemocionais.

Por uma questão de foco e simplicidade, recomendamos usar poucas métricas de resultado, e estas que sejam altamente relevantes para avaliar o impacto do projeto. Essas métricas devem estar intimamente ligadas aos resultados descritos na teoria da mudança (Passo 4).

Boas métricas possuem os seguintes atributos⁷:

- São altamente **relevantes** em termos de geração de impacto. Em outras palavras, elas devem representar resultados altamente desejáveis para as populações-alvo. Se a métrica segue a teoria da mudança descrita no Passo 4, e se a teoria identifica corretamente os principais resultados desejáveis, então a métrica deve ser relevante.
- São passíveis de ser **afetadas** pela ação dos executores do projeto. Deve-se evitar, por exemplo, indicadores que sejam muito agregados e amplos, cujo resultado seja afetado por uma miríade de fatores que estejam fora do controle direto dos executores. Em geral, métricas vinculadas aos resultados das atividades tendem a ser mais sujeitas à ação dos gerentes do projeto do que métricas relacionadas a transformações sociais mais amplas.
- Devem ser medidas e verificadas com suficiente **precisão**. Em geral, métricas baseadas em informações objetivas ("dados concretos") são preferíveis às baseadas em indicadores subjetivos, com alto potencial de erro de medição. Além disso, os dados não devem ser suscetíveis à manipulação por gestores, investidores e outras partes interessadas em relatar resultados positivos. Nesse sentido, dados coletados por terceiros (como institutos independentes) e amplamente utilizados por atores do setor de projetos tendem a ser preferíveis no que diz respeito ao critério de precisão.
- Devem ter **baixo custo** de medição, incluindo o custo de coleta de dados de populações usadas na comparação (ou seja, indivíduos no grupo de controle). O custo de medição tende a cair com o uso de dados publicamente disponíveis (secundários), em comparação a dados customizados (primários), coletados especificamente para o projeto.



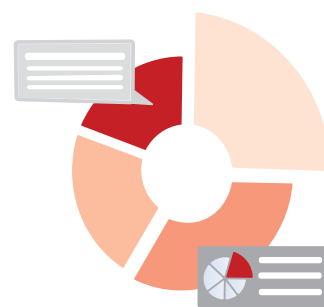
⁷ Essas características são particularmente relevantes, uma vez que a avaliação de impacto também pode ser usada para definir metas de gestão; ver: ROBERTS, J. Designing incentives in organizations. *Journal of Institutional Economics*, v. 6, n. 1, p. 125–132, mar. 2010. Características de boas métricas também são discutidas em: KUSEK, J. ZALL; RIST, R. **Ten Steps to a Results Based Monitoring and Evaluation System: A Handbook for Development Practitioners**. Washington: World Bank Publications, 2004.

Como sugestão, ao considerar métricas alternativas, cada uma delas deve ser avaliada com base nos critérios acima, selecionando-se as que melhor se alinharem à maioria dos atributos. Infelizmente, na maioria dos casos, não é possível atender a todos esses critérios. Os gestores devem então selecionar as métricas que melhor reflitam os objetivos do projeto, bem como aquelas amplamente utilizadas e validadas na atividade de interesse.

Considerando os atributos discutidos acima, o **Box 3** ilustra uma aplicação da ferramenta do **menu de métricas**: uma forma de comparar métricas alternativas, identificando visualmente suas vantagens e limitações. Para cada atributo, as métricas recebem classificações com base nas rubricas descritas no **Box 4**, em uma escala de 1 a 4.

Recomendamos que todas as métricas sejam comparadas com base em seus atributos individuais. No exemplo do **Box 3**, fica claro que todas as métricas têm vantagens e desvantagens. No entanto, como possibilidade de promover uma comparação geral de métricas, pode-se multiplicar as pontuações específicas de cada atributo, assumindo que a métrica preferível será aquela que pontuar bem na maioria dos atributos.

Por exemplo, a segunda métrica relacionada ao local de atividades ligadas ao turismo tem uma pontuação agregada (multiplicativa) de $4 \times 3 \times 4 \times 4 = 192$. A terceira métrica, que avalia a satisfação do usuário, tem, por sua vez, uma pontuação agregada de $4 \times 4 \times 1 \times 1 = 16$. Portanto, a primeira métrica é mais atraente do que a última. No entanto, isso não significa necessariamente que a satisfação do usuário deva ser ignorada. Se uma métrica estiver vinculada a uma saída da teoria da mudança, ela provavelmente será relevante e deverá ser considerada no painel de monitoramento. Assim, os gestores podem avaliar maneiras alternativas de aumentar a atratividade da métrica (por exemplo, reduzir os custos de coleta de dados) ou encontrar outras maneiras de medir o mesmo resultado.



Se uma métrica estiver vinculada a uma saída da teoria da mudança, ela provavelmente será relevante e deverá ser considerada no painel de monitoramento.

BOX 3. Menu de métricas: Exemplo de análise comparativa envolvendo métricas potenciais aplicadas a um parque turístico.

O exemplo a seguir ilustra a escolha de métricas potenciais para um parque em uma área de conservação florestal e com atrações turísticas. O governo tem interesse em estabelecer metas contratuais de impacto para nortear a ação de uma empresa privada selecionada para administrar o parque. Há uma preocupação particular em desenvolver os municípios localizados no entorno do local, além de garantir que a empresa busque a preservação do meio ambiente. Na análise abaixo, cada métrica é avaliada com base nas rubricas apresentadas no **Box 4**. A pontuação final é a multiplicação das avaliações individuais para cada atributo da métrica. As métricas relacionadas à renda são relevantes, precisas e de baixo custo; por outro lado, as métricas que mensuram os ganhos de renda das atividades relacionadas ao turismo são mais sensíveis às ações da empresa privada, que pode contratar trabalhadores locais para operar no parque ou estimular empresários da região a oferecer serviços complementares (como hotéis e restaurantes). A métrica de satisfação do usuário, embora relevante e sensível à ação, é cara e menos precisa – sendo baseada nas percepções dos usuários e sujeita a manipulação (por exemplo, os gestores podem influenciar a coleta de dados para que ocorra em momentos de baixa visitação, reduzindo assim a incidência de reclamações). Por fim, a métrica de aferição do percentual de áreas que seguem regulamentações ambientais, se informada pela administradora, pode obrigar o governo a contratar uma auditoria técnica externa, o que tende a aumentar os custos.

| RESULTADO (baseado na teoria da mudança) | MÉTRICA | RELEVANTE? | PASSÍVEL DE AÇÃO? | PRECISA? | DE BAIXO CUSTO? |
|--|---|---|---|--|---|
| Comunidades vizinhas obtêm maior renda e bem-estar. | Aumento da renda per capita dos municípios do parque e seu entorno. | ● O desenvolvimento local é um objetivo fundamental do projeto. | ○ A economia local pode ser afetada por outros fatores além das atividades do parque. | ● Há estatísticas confiáveis e publicamente disponíveis sobre a renda local. | ● No país onde o projeto está localizado, as estatísticas encontram-se disponíveis ao público e são facilmente acessíveis. |
| A comunidade local aumenta sua renda com atividades relacionadas ao turismo. | Aumento da renda per capita de serviços como hotéis, restaurantes, entre outros, nos municípios afetados pelo parque. | ● O desenvolvimento local é um objetivo fundamental do projeto. | ● O parque pode impulsionar a receita de hotéis e restaurantes, por exemplo, embora isso também dependa de outras condições contextuais. | ● Existem estatísticas confiáveis e publicamente disponíveis sobre a renda local. | ● As estatísticas estão disponíveis ao público e são divididas por setor da economia. |
| Os usuários do parque ficam mais satisfeitos. | Pontuação média de satisfação dos usuários, com base em pesquisas realizadas por institutos contratados. | ● Uma maior satisfação tende a refletir uma maior qualidade dos serviços do parque. | ● Um esforço para promover melhores serviços afeta diretamente a satisfação do usuário. | ○ Os indicadores de satisfação são mais subjetivos e podem estar sujeitos a manipulação. | ○ Para evitar a manipulação, será necessária uma coleta de dados dispendiosa por terceiros. |
| O parque atende aos regulamentos de conservação. | Porcentagem de áreas do parque em conformidade com as regulamentações ambientais. | ● Como o parque está localizado em uma área de preservação, a manutenção do meio ambiente é crítica. | ● Os gestores do parque podem controlar diretamente o status das áreas de conservação. | ○ Existem padrões claros de conservação, mas uma preocupação é que eles sejam monitorados pelos gestores do parque. | ○ Embora os gestores possam monitorar o status das áreas de conservação, pode ser necessária auditoria externa. |

BOX 4. Avaliação de métricas

As rubricas a seguir permitem a análise detalhada de cada métrica proposta de acordo com os quatro atributos: relevância, passível de ação, precisão e custo. Cada atributo recebe uma pontuação em uma escala de 1 a 4.

| | ○ 1 | ○ 2 | ● 3 | ● 4 |
|--------------------------|--|--|---|--|
| Relevante? | A métrica é pouco ou apenas tangencialmente ligada a algum resultado previsto na teoria da mudança. | A métrica é ligada a algum resultado previsto na teoria da mudança, mas não está claro como sua variação poderia ser valorizada pela população-alvo. | A métrica é ligada a algum resultado previsto na teoria da mudança, e sua variação será potencialmente valorizada pela população-alvo. | A métrica é ligada a algum resultado previsto na teoria da mudança, e sua variação será fortemente valorizada pela população-alvo. |
| Passível de ação? | Condições contextuais fora do alcance dos gestores envolvidos nas atividades previstas na teoria da mudança limitam substancialmente o possível efeito das atividades sobre a métrica. | A métrica pode possivelmente ser afetada por atividades previstas na teoria da mudança, porém esse efeito depende de diversas condições contextuais fora do alcance dos gestores envolvidos. | Existem argumentos lógicos que levam a crer que a métrica pode ser afetada por intervenções previstas na teoria da mudança. | Existem evidências de que a métrica pode ser fortemente afetada por intervenções previstas na teoria da mudança. |
| Precisa? | A métrica pouco reflete o resultado que se quer medir e/ou pode ser altamente manipulada e distorcida para erroneamente indicar um desempenho maior do que o efetivamente obtido. | A métrica reflete moderadamente o resultado que se quer medir e/ou envolve fontes de dados com considerável risco de manipulação ou cuja interpretação pode ser questionável. | A métrica reflete adequadamente o resultado que se quer medir e envolve fontes de dados menos reconhecidas, mas com baixo risco de manipulação. | A métrica reflete adequadamente o resultado que se quer medir e envolve fontes de dados reconhecidas e legitimadas. |
| De baixo custo? | A métrica requer medições específicas e custosas (por exemplo, <i>surveys</i> presenciais). | Existe um pequeno custo de coleta associado à métrica, devido à necessidade de medições específicas e/ou validação externa. | Os dados podem ser obtidos a baixo custo por meio de fontes existentes, mas cujo acesso depende de autorização ou esforço de terceiras partes. | Os dados podem ser obtidos a baixo custo por meio de fontes públicas ou de fácil acesso. |



6. DEFINIÇÃO DO MÉTODO PARA VERIFICAR ADICIONALIDADE



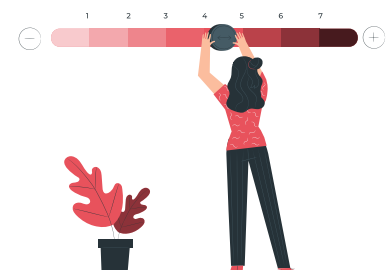
Depois de escolher as métricas que acompanharão os indicadores de desempenho operacional e, especialmente, os principais resultados, é importante definir como estas serão usadas para verificar a adicionalidade do projeto, ou seja, o que provavelmente teria acontecido com a população-alvo na ausência da intervenção. Esta etapa envolve a definição do grupo de tratamento (indivíduos e comunidades beneficiados pelo projeto) e do grupo de controle (indivíduos e comunidades que não receberam a intervenção), indicando o que teria acontecido aos tratados sem o projeto.

Como tende a ser difícil obter dados para ambos os grupos, de tratamento e de controle, recomendamos uma ênfase inicial em poucas métricas, como discutido na etapa anterior, que melhor atendam aos requisitos de relevância, precisão, possibilidade de ação e baixo custo. No exemplo do **Box 4**, a segunda métrica (receita de atividades relacionadas ao turismo) é uma candidata potencial para as próximas etapas envolvendo avaliação de impacto com base na adicionalidade, embora as outras métricas também possam ser seguidas para monitorar a evolução do projeto.

O método de verificação da adicionalidade varia dependendo do **nível de medição** escolhido para o projeto, conforme será detalhado na próxima seção. Níveis de medição mais altos aumentam a confiança de que as mudanças detectadas na população-alvo foram causadas pelo projeto; no entanto, níveis mais altos também tendem a aumentar a complexidade analítica e os custos. Nesta fase, principalmente se forem adotados níveis de medição mais elevados, recomendamos o apoio de especialistas para ajudar a identificar as técnicas mais adequadas, bem como as suas potenciais limitações.

Após escolher o método de verificação de adicionalidade, é importante definir com precisão, com base nas métricas selecionadas, como avaliar o impacto do projeto. No exemplo da **Figura 1**, o impacto é medido como sendo a diferença na variação da renda entre as comunidades tratada e de controle, considerando-se os dois períodos predefinidos. Essa medição nos permite verificar o **tamanho** total do impacto, definido como o ganho médio por indivíduo vezes o número total de indivíduos incluídos no grupo tratado.

Esta etapa também envolve uma definição do **horizonte de tempo** necessário para medir o impacto. Vários projetos exigem um horizonte temporal longo, devido às características inerentes à atividade e à necessidade de implementação de um conjunto complexo de ações complementares. Os passos 3 e 4, que envolvem benchmarking e definição da teoria da mudança, também possibilitam uma revisão de estudos anteriores, examinando não apenas o tamanho do impacto esperado, mas também quanto tempo foi necessário para que os resultados das intervenções ocorressem.



Após escolher o método de verificação de adicionalidade, é importante definir com precisão, com base nas métricas selecionadas, como avaliar o impacto do projeto.



7. PLANO DE AMOSTRAGEM



Definir o tamanho da amostra (o número de pessoas ou grupos que participarão do projeto) é mais uma etapa fundamental no plano de medição. Esta fase é particularmente importante caso sejam adotados níveis de medição mais elevados, utilizando técnicas que procuram medir com precisão o efeito causal da intervenção na população-alvo. Com um tamanho de amostra adequado, a diferença entre os resultados medidos antes e depois do projeto, bem como a diferença entre os grupos de tratamento e controle, pode ser medida com maior precisão estatística.

Essencialmente, uma amostra maior reduz o risco de que as análises não verifiquem um impacto positivo do projeto, supondo que ele realmente exista. Em geral, o tamanho da amostra recomendado depende do tamanho esperado do impacto. Tudo o mais constante, quanto menor o impacto esperado, maior será o tamanho da amostra necessária para fins estatísticos. No entanto, os custos de implementação do tratamento e da coleta dos dados relevantes também devem ser levados em consideração. Como o cálculo de tamanhos ótimos de amostra requer conhecimento prévio de estatística, é importante recorrer à literatura técnica existente e contar com as contribuições de especialistas em avaliação de impacto⁸.



8. CRONOGRAMA DE MEDIÇÃO



Nesta etapa, devem-se definir os períodos de coleta de dados, indicando claramente, se for o caso, as medições que ocorrerão antes e depois da intervenção (vide **Figura 1**). Os gestores também podem definir medições intermediárias como forma de acompanhar a evolução do projeto. Conforme discutido anteriormente, eles devem estar cientes de que, na maioria dos casos, é necessário um prazo razoável, após as intervenções esperadas, até que o impacto passe a ser observável. Por exemplo, em projetos relacionados a educação, em que a aprendizagem dos alunos requer um esforço contínuo e de longo prazo, o horizonte de tempo tende a ser longo, especialmente se houver interesse em medir os efeitos da conclusão de um ciclo educacional ou seu impacto sobre os ganhos futuros dos alunos.

A decisão sobre a janela temporal ideal para acompanhar e medir os resultados do projeto pode ser baseada em estudos anteriores que avaliaram o impacto de intervenções similares ao longo do tempo (conforme discutido no **Passo 3**, de *benchmarking*).

⁸ Uma discussão técnica sobre como definir tamanhos de amostra pode ser encontrada em: DUFLO, E.; GLENNERSTER, R.; KREMER, M. Using Randomization in Development Economics Research: A Toolkit. In: **Handbook of Development Economics**. Amsterdam: Elsevier, 2008. v. 4, p. 3895–3962. Uma ferramenta prática para calcular tamanhos mínimos de amostras pode ser encontrada em: RAUDENBUSH, S. W. et al. **Optimal Design Software for Multi-level and Longitudinal Research**, 2011. Disponível em: <https://sites.google.com/site/optimaldesignsoftware/home>. Acesso em: 20 mar. 2022.

Níveis de medição



Os níveis de medição propõem, essencialmente, formas alternativas de corrigir os efeitos de fatores externos decorrentes de diferenças preexistentes entre os grupos de tratamento e de controle. Por exemplo, suponha que um investidor de projetos educacionais queira avaliar os efeitos de um novo método de ensino e o ofereça a várias escolas. É possível que as escolas que aceitem aderir ao programa sejam justamente aquelas cujos gestores estejam mais motivados a buscar melhorias no aprendizado. Se essa variável afetando a escolha (“motivação gerencial”) não for observada pelo avaliador, ele poderá concluir erroneamente que as melhorias detectadas nas escolas que aderiram voluntariamente ao programa vieram das intervenções do programa em si, ainda que a causa real seja a propensão natural das escolas escolhidas para melhorar seu desempenho.

Para mitigar esse erro, o ideal é que os grupos de tratamento e de controle sejam escolhidos aleatoriamente. Com um número suficiente de casos (sorteios aleatórios), os grupos tendem a ser semelhantes em termos de atributos que o avaliador de impacto pode medir diretamente (como características demográficas), bem como atributos que são mais difíceis de observar diretamente (no exemplo acima, a motivação gerencial). Tal semelhança é fundamental porque, como discutido anteriormente, o objetivo do grupo de controle é indicar o que teria acontecido ao grupo de tratamento sem a intervenção. Em outras palavras, o grupo de controle deveria espelhar uma “realidade paralela” em que os indivíduos do grupo tratado não estivessem sujeitos ao projeto proposto.

No entanto, aleatorizar indivíduos ou grupos é um processo particularmente complexo e, em muitos casos, até mesmo inviável. Independentemente de os patrocinadores serem fundos, indivíduos, empresas ou governos, a decisão de investir ou não em um determinado projeto muitas vezes segue um plano pré-especificado. Portanto, na maioria dos casos, os patrocinadores e gestores de projeto precisam recorrer a métodos que tentem corrigir as diferenças preexistentes entre os grupos de tratamento e de controle.

Seja qual for o método de verificação de adicionalidade, é importante prover total **transparência** em relação às prováveis limitações do método. Especificamente, propomos abaixo **níveis de medição** alternativos que variam de acordo com a robustez desejada da avaliação, em termos do rigor com que o efeito *causal* do projeto é avaliado.

Conforme visto na **Figura 3**, iniciamos com um tipo de medição denominada *básica*, que não envolve avaliação de adicionalidade, como forma de mostrar como cada nível adicional permite maior rigor na estimativa de impacto. Essa abordagem é geralmente empregada em painéis de monitoramento que simplesmente desejem rastrear mudanças no grupo de tratamento (ou seja, aquele diretamente afetado pelo projeto). Como os gestores de projeto geralmente adotam mais de uma métrica, é possível que os vários indicadores selecionados sejam avaliados segundo níveis de medição distintos. Assim, é provável que a maioria dos resultados seja avaliada usando a abordagem básica, enquanto alguns resultados selecionados serão avaliados usando níveis de medição mais robustos.

Figura 3 – Níveis de medição para verificação de adicionalidade.

| | |
|----------------|--|
| NÍVEL 3 | Comparação do grupo de tratamento com um grupo de controle, no qual os tratados são escolhidos aleatoriamente (<i>Randomized Controlled Trial</i> ou RCT) |
| NÍVEL 2 | Comparação do grupo de tratamento com um grupo de controle considerado semelhante |
| NÍVEL 1 | Comparação do grupo de tratamento com dados agregados locais ou regionais |
| BÁSICO | Sem verificação de adicionalidade; apenas a evolução do grupo de tratamento é avaliada ao longo do tempo |

BÁSICO – MEDIÇÃO DA EVOLUÇÃO DO GRUPO DE TRATAMENTO

Esta abordagem *não envolve* verificação de adicionalidade e *não deve* ser considerada um nível de medição, uma vez que não adota análise contrafactual (ou seja, avaliação do que provavelmente teria acontecido com a população-alvo sem o projeto). Porém, para uma ampla gama de métricas, é o método mais viável, visto que oferece maior facilidade de coletar dados de indivíduos e grupos sujeitos à intervenção. Nesse caso, simplesmente observamos como certas métricas variaram ao longo do tempo, antes e depois da data de início do projeto. Essa prática também é amplamente usada para monitorar um *scorecard* de métricas de desempenho específicas do projeto para as populações tratadas, conforme discutido no Passo 4. Apesar de sua simplicidade e praticidade, os gestores devem evitar inferir automaticamente que qualquer variação positiva nos resultados medidos foi causada diretamente pelo projeto. A própria população-alvo pode já ter uma tendência de melhoria antes mesmo do início das intervenções. Além disso, o projeto pode ter se beneficiado de mudanças externas que afetaram positivamente a população-alvo. Portanto, os gestores que adotam a abordagem básica devem tratar as medições simplesmente como indicadores de mudanças no grupo de tratamento, as quais podem ou não ter sido causadas pelas intervenções propostas pelo projeto.

NÍVEL 1 – MEDIÇÃO COMPARANDO O GRUPO DE TRATAMENTO COM OS DADOS AGREGADOS LOCAIS OU REGIONAIS

Neste nível, usamos dados agregados como uma forma de comparação entre os resultados do grupo de tratamento e o que teria acontecido a esse grupo sem a intervenção. Para fins de comparação, é recomendável usar dados agregados que já estejam disponíveis para a região ou o local onde o investimento está sendo feito. É importante obter dados sobre o que aconteceu com o alvo em comparação às variações nas métricas agregadas, antes e depois da intervenção (**Box 5**).

No presente nível, os gestores também devem levar em consideração o tamanho da intervenção. Se o projeto afetar um grande número de indivíduos em uma determinada região, os dados agregados serão significativamente afetados pela intervenção, e a comparação não será apropriada. Uma alternativa, nesse caso, é comparar o grupo de tratamento com indivíduos em outras localidades próximas e semelhantes.

BOX 5. Exemplo de verificação de adicionalidade no Nível 1

Uma empresa de processamento de alimentos decide adquirir produtos agrícolas de pequenos agricultores familiares em alguns municípios de baixa renda, ao mesmo tempo em que oferece apoio técnico e gerencial a esses agricultores. A definição das comunidades-alvo é deliberadamente escolhida pelo gerente do projeto. Usando a renda dos agricultores como métrica de resultado, a avaliação do impacto é feita comparando-se a renda das famílias apoiadas pelo projeto com a renda agregada das famílias nas áreas rurais do município ou da localidade onde o investimento foi feito, antes e depois da intervenção. A adicionalidade no Nível 1 será verificada se a variação da renda das famílias apoiadas for maior do que a variação da renda agregada das famílias da mesma região.

NÍVEL 2 – MEDIÇÃO USANDO UM GRUPO DE CONTROLE COM CARACTERÍSTICAS SEMELHANTES ÀS DO GRUPO DE TRATAMENTO

Neste nível, embora a aleatorização não tenha sido realizada, busca-se criar um grupo de controle com indivíduos ou comunidades comparáveis aos casos tratados. Por esse motivo, a verificação da adicionalidade no Nível 2 envolve técnicas que são chamadas de **não experimentais**. Assim como no Nível 1, quem virá a ser beneficiado com o projeto já está predeterminado. Além disso, no processo de avaliação, são implementadas técnicas para a construção de um grupo-controle de indivíduos semelhantes aos do grupo de tratamento. Em outras palavras, em vez de usar dados agregados de um determinado local, procura-se rastrear grupos ou indivíduos que são semelhantes aos indivíduos tratados, mas que não receberam o investimento.

Uma maneira comum de avaliar o impacto no Nível 2 é empregar o método de **diferenças-em-diferenças**, exemplificado na **Figura 1**. Seguindo esse método, o impacto do projeto é medido como sendo a diferença entre a evolução das métricas de resultado do grupo de tratamento e a evolução das métricas de resultado do grupo de controle. A evolução do grupo de controle serve essencialmente como uma estimativa contrafactual: o que teria acontecido à população-alvo sem a intervenção

Observe que o Nível 2 requer que os grupos de tratamento e de controle sejam realmente comparáveis em termos de suas características subjacentes e da evolução de suas métricas de resultado. Assim, no método de diferenças-em-diferenças, as métricas de resultados dos grupos de tratamento e de controle deveriam exibir uma **tendência** semelhante após a intervenção, caso ela não tivesse ocorrido. No entanto, como não se pode testar se as tendências são paralelas após a intervenção, os pesquisadores geralmente testam essa suposição observando as tendências anteriores. A ideia é que, se as tendências anteriores ao tratamento apresentavam evoluções paralelas para os grupos tratados e não tratados, seria possível esperar que elas continuariam paralelas após a implementação do tratamento, caso este não houvesse sido implementado. Na Figura 1, as comunidades tratadas e de controle evoluíram "em paralelo" antes do início do projeto, o que sugere que foram afetadas por fatores semelhantes. Tendências não paralelas poderiam indicar que os grupos estariam evoluindo de maneiras distintas, devido a vários fatores não relacionados ao projeto.



Outra forma comum de escolha de indivíduos com características semelhantes entre os grupos de controle e de tratamento envolve uma técnica chamada **pareamento**. Nessa técnica, com base em características observáveis que são relevantes para influenciar a participação dos indivíduos no programa (como idade, sexo, renda, etc.), procura-se encontrar uma ou mais unidade(s) no grupo de controle que sejam as mais parecidas possíveis com cada unidade no grupo de tratamento (ver exemplo 6.1. no **Box 6**). A técnica de pareamento também é comumente combinada com a técnica de diferenças-em-diferenças.

Nos casos em que haja competição para participar no projeto e um determinado critério de seleção, uma possibilidade é considerar a **descontinuidade** que surge quando observamos participantes próximos a um determinado patamar: aqueles que quase foram rejeitados (tratados) e os que quase foram aceitos no projeto, mas ficaram abaixo do limite de seleção (controle) (ver exemplo 6.2 no **Box 6**).

Ainda em outras situações em que apenas um caso particular é avaliado (por exemplo, uma única organização ou um só município), existe a possibilidade de se criar um **controle sintético** a partir da combinação de outros casos que coletivamente se aproximem do caso tratado. Essa técnica requer dados do caso tratado e de outros casos potenciais antes do início do projeto, de forma a criar um controle sintético que possa replicar de forma adequada a evolução anterior do caso que recebeu a intervenção⁹.

Ao invés de usar dados agregados de um determinado local, procura-se rastrear grupos ou indivíduos que são semelhantes aos indivíduos tratados, mas que não receberam o investimento.

BOX 6. Exemplos de verificação de adicionalidade no Nível 2

Exemplo 6.1: Em um projeto no qual o resultado a ser avaliado é a renda familiar (ver **Box 3**), comparamos os níveis de renda das famílias tratadas com os dados das famílias do grupo de controle. O gestor do projeto define deliberadamente quem será beneficiado com a intervenção. As famílias no grupo de controle são então selecionadas com base em certas características observáveis, que são semelhantes às do grupo tratado. Por exemplo, um grupo de controle pode consistir em um conjunto de famílias com níveis iniciais de renda, escolaridade e número de filhos semelhantes aos níveis observados nas famílias tratadas.

Exemplo 6.2: Uma organização deseja implementar um projeto de microcrédito. Os gestores podem convidar pequenos empreendedores para participar do programa e, em seguida, classificar esses empreendedores com base em indicadores de risco de crédito (receita, histórico etc.). Se houver dinheiro limitado a ser desembolsado com o microcrédito, apenas os mais bem classificados serão selecionados, com base em um critério de corte que define um risco de crédito mínimo aceitável. Na descontinuidade desse corte específico, podemos supor que os indivíduos são mais ou menos semelhantes entre si, uma vez que possuem avaliações de risco de crédito semelhantes. Assim, podemos comparar o grupo de empreendedores selecionados logo acima do ponto de corte com aqueles que não foram selecionados, mas estavam logo abaixo do ponto de corte de seleção no *ranking*.

⁹ Como exemplo da técnica de pareamento, ver: ABADIE, A. et al. Implementing Matching Estimators for Average Treatment Effects in Stata. *The Stata Journal: Promoting communications on statistics and Stata*, v. 4, n. 3, p. 290–311, 1 ago. 2004. Para uma discussão geral envolvendo vários métodos, ver: ANGRIST, J. D.; PISCHKE, J. S. *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press, 2009; e: MENEZES FILHO, N. (ORG.). *Avaliação Econômica de Projetos Sociais*. São Paulo: Dinâmica Gráfica e Editora Ltda, 2012. A técnica de controle sintético pode ser vista em: ABADIE, A.; DIAMOND, A.; HAINMUELLER, A. J. Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program. *Journal of the American Statistical Association*, v. 105, n. 490, p. 493–505, jun. 2010.

NÍVEL 3 – MENSURAÇÃO COM SORTEIO (*Randomized Controlled Trial RCT*)

O Nível 3, que envolve técnicas **experimentais**, permite máxima confiabilidade na estimativa de avaliação de impacto. Nesse nível, a decisão de quais indivíduos ou grupos estarão sujeitos à intervenção é feita de forma aleatória, reduzindo, assim, potenciais vieses na seleção dos beneficiários do projeto. Nos casos em que não seja possível nem desejável excluir determinados grupos, uma alternativa é implementar um desenho experimental em que o grupo selecionado aleatoriamente seja encorajado a adotar o que é proposto pela intervenção (**Box 7**).

Se realizada corretamente, a seleção aleatória garantirá que não haja diferenças entre os grupos de tratamento e de controle. Por esse motivo, as medições do Nível 3, em geral, não requerem avaliar os resultados antes da intervenção, conforme indicado na **Figura 1**. Seguindo os procedimentos de aleatorização adequados, os avaliadores de impacto podem medir o desempenho comparativo dos grupos após a implementação do projeto.

Neste nível, é particularmente importante prestar atenção aos cálculos corretos do tamanho mínimo da amostra, conforme discutido na Etapa 7 da seção anterior. Também é importante distinguir entre amostragem aleatória e seleção aleatória de grupos tratados. Por exemplo, suponha que um determinado projeto educacional visasse 30 escolas de tal forma que a escolha das escolas fosse arbitrária, sem aleatorização. O Nível 3 não seria válido nesse caso específico, mesmo se os gestores escolhessem aleatoriamente outras 30 escolas para servir como grupo de controle. Isso ocorre porque a seleção daqueles que receberam a intervenção não foi baseada na atribuição aleatória, criando assim o risco de que outros fatores não observáveis viessem a afetar a avaliação de impacto do grupo tratado.



¹⁰ Para avaliar o efeito da intervenção neste caso, uma opção é usar o estimador conhecido como LATE (Local Average Treatment Effect). Ver: ANGRIST, J. D.; PISCHKE, J. S. **Mostly harmless econometrics: An empiricist's companion**. Princeton: Princeton University Press, 2009.

BOX 7. Exemplos de verificação de adicionalidade no Nível 3

Exemplo 7.1: Uma empresa oferece um programa de microcrédito em um local onde não há possibilidade de financiar a comunidade como um todo. Os gestores podem convidar destinatários em potencial e, em seguida, escolher aleatoriamente quem receberá o crédito. Alternativamente, a empresa pode emprestar para os empreendedores mais bem classificados em termos de risco de crédito e, em seguida, aleatorizar entre o conjunto de pessoas que foram quase selecionadas, ou seja, que estavam logo abaixo do risco de crédito mínimo adotado como critério de corte. Tomando esse conjunto de empreendedores, a empresa pode escolher aleatoriamente um grupo de indivíduos tratados que receberão crédito; o grupo de indivíduos não selecionados segundo esse sorteio passará a constituir o grupo de controle.

Exemplo 7.2: Os alunos de um amplo conjunto de escolas públicas têm acesso a uma plataforma tecnológica de aprendizagem online. Embora todos os alunos tenham acesso à plataforma, um grupo de alunos é selecionado aleatoriamente para receber mensagens de estímulo ao seu uso. Considerando que apenas uma fração dos alunos que receberam as mensagens irão de fato utilizar a plataforma (ou seja, apenas aqueles alunos serão efetivamente atendidos), é possível comparar o desempenho desses alunos com o desempenho do grupo de controle de alunos que não receberam as mensagens. Essa análise, entretanto, requer ajustes estatísticos, uma vez que deve ser avaliado o desempenho de quem efetivamente utilizou a plataforma, não necessariamente de quem recebeu ou não o estímulo¹⁰.

COMPARANDO OS NÍVEIS DE MEDIÇÃO

A Tabela 1 apresenta uma avaliação comparativa dos benefícios e limitações de cada nível de medição. Em geral, a avaliação do impacto se torna mais robusta à medida que avançamos para o Nível 3, havendo maior confiança de que o projeto de fato causou os resultados detectados na população-alvo. Porém, ao mesmo tempo, aumenta-se a complexidade do desenho e das análises, exigindo-se mais recursos financeiros e técnicos para realizar a avaliação.

Tabela 1 – Vantagens e limitações dos níveis de medição

| | VANTAGENS | LIMITAÇÕES |
|----------------|---|--|
| BÁSICO | <ul style="list-style-type: none">· Dados mais facilmente coletados (do próprio projeto e de sua população-alvo).· Maior simplicidade nas análises (comparações simples do que aconteceu com a população-alvo antes e depois do projeto). | <ul style="list-style-type: none">· Como não há avaliação contrafactual (o que teria acontecido com a população-alvo sem o projeto), não há verificação de adicionalidade. |
| NÍVEL 1 | <ul style="list-style-type: none">· Na maioria dos casos, os gestores têm acesso apenas a dados agregados (por exemplo, renda municipal em vez da renda das famílias do município).· Comparado a outros métodos de verificação de adicionalidade, a análise é mais simples (comparação dos resultados do projeto com os resultados agregados). | <ul style="list-style-type: none">· As características da população agregada podem ser muito distintas daquelas da população-alvo, ainda que dentro do mesmo recorte geográfico.· A comparação não permite uma avaliação de efeito causal do projeto, uma vez que não há uso de técnicas para controlar as potenciais diferenças entre os grupos. |
| NÍVEL 2 | <ul style="list-style-type: none">· Na maioria dos casos, a população-alvo já está definida, inviabilizando a aleatorização.· Ao criar grupos de controle com características semelhantes às da população-alvo, os métodos de Nível 2 evitam o viés gerado por diferenças substanciais entre os grupos.· As várias técnicas do Nível 2 permitem inferência estatística sobre o impacto observado no grupo tratado, frente ao grupo de controle. | <ul style="list-style-type: none">· As técnicas de Nível 2 só identificam o efeito causal da intervenção se assumirmos que não há influência crítica de fatores que não são observados e medidos – um pressuposto raramente verificado na prática.· É necessário reunir informações sobre as características dos indivíduos incluídos tanto no grupo de tratamento quanto no de controle.· As técnicas de Nível 2 requerem a existência de casos no grupo de controle que sejam próximos, em suas características subjacentes, aos casos no grupo de controle. Caso contrário, não é possível realizar a comparação. |
| NÍVEL 3 | <ul style="list-style-type: none">· Por meio da aleatorização, é possível garantir com maior segurança que o efeito medido foi efetivamente causado pelo projeto.· Em geral, a aleatorização não requer a medição dos resultados antes do início do projeto; os sujeitos podem ser comparados após a implementação das intervenções. | <ul style="list-style-type: none">· As técnicas do Nível 3 exigem atenção especial ao tamanho da amostra. Seu uso pode ser inviável quando houver poucos casos recebendo a intervenção.· Projetos experimentais do Nível 3 estão altamente sujeitos a efeitos cruzados, atrito e outros problemas discutidos na próxima seção do <i>Guia</i>.· A aleatorização pode criar dilemas éticos nos casos em que seja problemático excluir indivíduos de uma determinada intervenção. |

CUIDADOS DURANTE A MEDIÇÃO

Algumas precauções são extremamente relevantes para que se possa realizar uma avaliação de impacto de maneira satisfatória. Abaixo, são descritos alguns dos problemas críticos que o executor do projeto deve considerar ao medir seu impacto¹¹.

EXTERNALIDADES

Externalidades ou “efeitos cruzados” entre indivíduos de grupos diferentes devem ser minimizadas tanto quanto possível. Considere, por exemplo, que se queira avaliar o impacto de uma iniciativa focada no aumento da renda de uma determinada localidade. Para isso, antes do início do projeto, são coletados dados dessa comunidade e da comunidade do entorno, que não é beneficiada pelo projeto. A externalidade ocorre se a comunidade tratada, que recebe uma renda maior como resultado do investimento, puder gastar parte de seus recursos na compra de produtos ou na transferência de renda para a outra comunidade. Caso isso ocorra, o impacto adicional do programa será subestimado. Esse tipo de externalidades é problemático para avaliações de impacto, pois pode distorcer o resultado ao “contaminar” o grupo de controle com a intervenção focada no grupo tratado.

Para contornar esse problema, é importante garantir que não haja comunicação, troca de informações ou troca física de bens e serviços entre os diferentes grupos. Uma forma de proceder é certificar-se de que os grupos de tratamento e de controle estejam geograficamente distantes ou isolados, diminuindo assim o risco de efeitos cruzados.

ATRITO

Dado o tempo entre a intervenção e a avaliação subsequente, é possível que ocorra um “descasamento” entre os indivíduos inicialmente observados e os monitorados na rodada final de medição, tanto no grupo de controle quanto no grupo de tratamento. Por exemplo, alguns indivíduos que foram observados no momento inicial podem não mais residir no mesmo local durante a rodada final de medição.

Vale notar que sorteios não evitam completamente o atrito não aleatório ao final da medição. Além disso, os problemas de atrito reforçam a necessidade de definir um tamanho de amostra suficientemente grande, antes de realizar a medição. Um exemplo prático desse problema pode ser visto na avaliação de projetos que visam aumentar a renda em comunidades rurais. Mesmo que os grupos inicialmente possam ter sido selecionados de forma aleatória, a saída dos indivíduos que não receberam a transferência de renda frente à continuidade dos indivíduos que receberam a transferência pode causar um



Vale notar que sorteios não evitam completamente o atrito não aleatório ao final da medição.

¹¹ Esta seção é fortemente baseada em: DUFLO, E.; GLENNERSTER, R.; KREMER, M. Using Randomization in Development Economics Research: A Toolkit. In: *Handbook of Development Economics*. Amsterdam: Elsevier, 2008. v. 4, p. 3895–3962

atrito não aleatório. Nesse caso, o atrito produz vieses, pois ocorre uma perda não aleatória de uma grande quantidade de dados dos indivíduos do grupo de controle.

É sempre importante ser transparente sobre o nível de atrito, tanto no grupo de tratamento quanto no grupo de controle. Em outras palavras, é necessário informar quantos indivíduos nos grupos de tratamento e de controle foram monitorados inicialmente, mas não foram encontrados nas rodadas subsequentes de medição. Também é desejável comparar as características médias desses grupos antes e depois do atrito.

EFEITO HAWTHORNE E EFEITO JOHN HENRY

Finalmente, outra limitação importante na medição do impacto é a mudança comportamental dos indivíduos tanto do grupo de tratamento, quanto do grupo de controle. O chamado **efeito Hawthorne** pode ocorrer quando os indivíduos do grupo de tratamento percebem que estão sob alguma forma de intervenção e, conseqüentemente, mudam seu comportamento de acordo com esse fato¹². Em projetos de educação, por exemplo, professores ou alunos podem aumentar seus esforços em função da consciência de que o grupo está sendo beneficiado pela intervenção.

O **efeito John Henry**, por sua vez, pode ocorrer quando professores do grupo de controle se sentem desafiados e passam a competir com seus colegas do grupo de tratamento para mostrar que também seriam elegíveis para se beneficiar do mesmo programa¹³. Por outro lado, se os professores ficarem menos motivados por não receber os benefícios do grupo tratado, o impacto do programa será superestimado. Em ambos os casos, o efeito real do programa pode ser distorcido.

Embora as respostas comportamentais sejam sempre complexas, há maneiras de projetar a medição para minimizar os efeitos Hawthorne e John Henry. Por exemplo, três grupos podem ser usados: tratamento, controle e **placebo**. Este último é um grupo que é observado ou interage brevemente com os executores do projeto, mas não recebe o conjunto completo de intervenções. Considere, por exemplo, um projeto em que uma grande empresa deseje melhorar a renda de certas comunidades adquirindo seus insumos locais. Nesse caso, a empresa deve não apenas considerar uma comunidade de fornecedores reais (tratamento) e uma comunidade comparável de não fornecedores (controle), mas também monitorar uma comunidade placebo de não fornecedores com os quais possua alguma forma de relacionamento (por exemplo, a empresa poderia enviar gerentes de aquisição para contatar as comunidades sem realmente assinar contratos de aquisição). A evidência de impacto aumenta quando o grupo tratado exhibe melhorias acima e além dos grupos controle e placebo.

O chamado efeito Hawthorne pode ocorrer quando os indivíduos do grupo de tratamento percebem que estão sob alguma forma de intervenção e, conseqüentemente, mudam seu comportamento de acordo com esse fato.

¹² O efeito Hawthorne foi batizado em homenagem a uma experiência ocorrida em 1927 na fábrica da West Electric Company (Chicago, EUA), onde se constatou que os indivíduos alteravam seu comportamento de acordo com as mudanças percebidas no ambiente de trabalho.

¹³ O efeito John Henry é baseado na história de um trabalhador em um local onde foi introduzida uma nova ferramenta para aumentar a produtividade, ainda em fase de testes. Nessa ocasião, a pessoa se sentiu desafiada a trabalhar mais para mostrar que poderia superar a nova tecnologia.

Monetização do impacto do projeto

O presente *Guia* enfatiza os procedimentos para estimar o impacto de um projeto com base em certos objetivos e resultados delineados pela teoria da mudança proposta. Portanto, não é objetivo do *Guia* discutir como computar o retorno econômico da intervenção socioambiental, em função do quanto foi gasto para promovê-la, em comparação com os benefícios gerados aos indivíduos tratados. No entanto, nesta seção final, fornecemos algumas indicações breves de como avaliar o impacto do projeto do ponto de vista econômico.

Quando houver dados de custo disponíveis sobre as intervenções, os gestores podem começar computando indicadores de **custo-efetividade**. Por exemplo, suponha que um programa de treinamento voltado para 1.000 pessoas desempregadas conseguiu empregar 100 pessoas adicionais em comparação ao que teria acontecido sem o projeto. Como visto nas seções anteriores, essa estimativa de impacto (adicionalidade) pode vir de comparações entre os grupos de tratamento e de controle. Se o programa de treinamento custar \$200.000, sua relação custo-efetividade será de $\$200.000/100 = \2.000 para cada indivíduo *adicional* que encontrar um emprego.

Outra possibilidade é monetizar os resultados avaliados. A relação **custo-benefício** do projeto indica quanto impacto foi gerado (em termos monetários) para cada unidade monetária investida no grupo de tratamento. Por exemplo, um estudo calculou os benefícios monetários de investir em melhorias de infraestrutura em saneamento – incluindo tratamento de saúde evitado devido à menor incidência de doenças e ao tempo extra que indivíduos mais saudáveis podem gastar em suas atividades produtivas. A relação custo-benefício estimada foi de 5,5: cada unidade monetária investida no projeto deve gerar 5,5 unidades monetárias em benefícios sociais¹⁴.

Em alguns casos, com essas proporções, pode-se comparar os resultados sociais ou ambientais de dois ou mais projetos distintos. Por exemplo, um estudo em educação mostrou que aumentar a carga horária escolar de 4 para 5 horas ou reduzir o tamanho das turmas de 38 para 30 alunos teve impactos semelhantes nas notas dos estudantes. Ainda assim, quando seus custos foram considerados, o projeto envolvendo aumento da jornada escolar teve a maior relação custo-benefício¹⁵.

Existem também várias maneiras de monetizar os resultados. Por exemplo, com base no exemplo do **Box 2**, um projeto educacional em escolas de segundo grau pode melhorar o aprendizado dos alunos e aumentar a probabilidade de eles irem para uma boa universidade, com impacto positivo em seus ganhos de renda futuros. Esses ganhos podem ser comparados ao custo do projeto.

Outra possibilidade é estimar a disposição de pagar dos beneficiários. A metodologia envolve perguntar diretamente ao público de interesse quanto ele estaria disposto a pagar por um determinado bem ou serviço. Às vezes, esse método é usado para valorar projetos ou iniciativas culturais destinadas a aumentar a satisfação com uma determinada atividade (como visitas a parques ou museus). O “preço” informado pelos usuários do serviço pode então ser comparado ao seu custo.

Para mais detalhes sobre como monetizar impacto, consulte a publicação do Insper Metricis:

[Monetização de Impacto Social.](#)

¹⁴ Ver: HUTTON, G. Global costs and benefits of reaching universal coverage of sanitation and drinking-water supply. *Journal of Water and Health*, v. 11, n. 1, p. 1–12, 2013. Para uma discussão mais geral sobre as análises de custo-efetividade e custo-benefício, ver: BOARDMAN, A. E. et al. *Cost-benefit analysis: concepts and practice*. Cambridge: Cambridge, 2018.

¹⁵ Ver: OLIVEIRA, J. M. DE. Custo-efetividade de políticas de redução do tamanho da classe e ampliação da jornada escolar: uma aplicação de estimadores de matching. *Revista do BNDES*, n. 33, 2010. Disponível em: https://web.bndes.gov.br/bib/jspui/bitstream/1408/2567/1/Custo%20Efetividade%20de%20Políticas%20de%20Reducao%20do%20Tamanho%20da%20Classe%20e%20Ampliao%20da%20Jornada%20Escolar_P_BD.pdf. Acessado em: 23 mar. 2022.

Dados Internacionais de Catalogação na Publicação (CIP)

159g Insper Metricis

Guia de avaliação de impacto socioambiental para utilização em projetos e investimentos de impacto: guia geral com foco em monitoramento e verificação de adicionalidade. - 5a. ed. - São Paulo: Insper, 2022.

30 p. ill.: col.

ISBN 978-65-991497-4-0

1. Investimentos de impacto 2. Avaliação de impacto
3. Verificação de adicionalidade (additionality)
4. Socioambiental 5. Gestão de projetos

I. Autor II. Título

CDU 504

Catalogação na Fonte: Ricardo Rodrigues Ramos CRB 8/9309

Insper METRICiS

Núcleo de Medição para Investimentos de Impacto Socioambiental