



Guia de Avaliação de Impacto Socioambiental para Utilização em Negócios e Investimentos de Impacto

GUIA GERAL COM FOCO EM VERIFICAÇÃO DE ADICIONALIDADE

Insper METRICiS

Núcleo de Medição para Investimentos de Impacto Socioambiental

São Paulo

Quarta revisão, 2 de abril de 2018

SOBRE O INSPER METRICIS

O Núcleo de Medição para Investimentos de Impacto Socioambiental do Insper realiza estudos sobre investimentos e estratégias empresariais de alto impacto socioambiental. Ênfase especial é dada ao desenvolvimento de ferramentas para avaliação e remuneração por impacto, bem como para a validação de medidas realizadas por organizações e governos.

Com procedimentos de avaliação claros e objetivos, validados por uma organização independente, é possível estimular não somente o crescimento dos investimentos e negócios de impacto, como também impulsionar novas formas de captação e financiamento destes projetos.

Além disso, com o aprendizado gerado pelos projetos, espera-se um contínuo registro e aprimoramento de melhores práticas por meio de casos e pesquisas ligados aos projetos de alto impacto.

SOBRE O INSPER

O Insper é uma instituição independente e sem fins lucrativos, dedicada ao ensino e à pesquisa nas áreas de Administração, Economia, Direito e Engenharia. Tem como missão ser um centro de referência, explorando as complementariedades dessas áreas.

Suas atividades de ensino abrangem cursos para várias etapas de uma trajetória profissional: graduação (Administração, Economia e Engenharias), pós-graduação lato e stricto sensu (Certificates, MBAs, programas da área de Direito, Mestrados Profissionais e Doutorado) e Educação Executiva (programas de curta e média duração, e customizados de acordo com as necessidades das empresas).

No âmbito da produção de conhecimento, a instituição atua por meio cátedras e centros de pesquisa que reúnem pesquisadores em estudos e projetos dirigidos a políticas públicas (CPP), finanças (CeFi) e negócios (CENeg). A instituição ainda conta com centros que promovem o empreendedorismo (CEMP) e a liderança e inovação (CLI). Tem as certificações de qualidade da AACSB, AMBA e Anamba. O Insper Metricis é um núcleo do Centro de Estudos em Negócios (CENeg).

GUIA DE AVALIAÇÃO DE IMPACTO SOCIOAMBIENTAL PARA UTILIZAÇÃO EM NEGÓCIOS E INVESTIMENTOS DE IMPACTO

Uma versão inicial desse documento foi elaborada por Sérgio G. Lazzarini, Leandro S. Pongeluppe, Pui Shen Yoong e Nobuiuki Costa Ito. Uma nova revisão contou com as contribuições de Mariana Suplicy, Rafael Vivolo, Guilherme Lichand, Amanda Arabage, Carlos Kazunari Takahashi, Sandro Cabral e José Geraldo Setter Filho. O texto se beneficiou de sugestões preliminares de grupo técnico de discussão envolvendo as seguintes pessoas: Angélica Rontondaro, Célia Cruz, Franco Veludo, Frederik Kuonen, Simon Locher, Tatiana Fonseca, Rafaella Ziegert e Raquel Costa. Agradecemos o apoio financeiro do Latin America Impact Economy Innovations Fund (Rockefeller Foundation, Avina e Omidyar), em proposta coordenada pelo ICE (Célia Cruz e Maria Amélia Sampaio). Agradecemos também as sugestões de Fernando Carnaúba, Ben Carpenter, Luis Fernando Guedes Pinto, Naercio Menezes Filho, Ricardo Paes de Barros, Brian Trelstad e Maurício Voivodic.

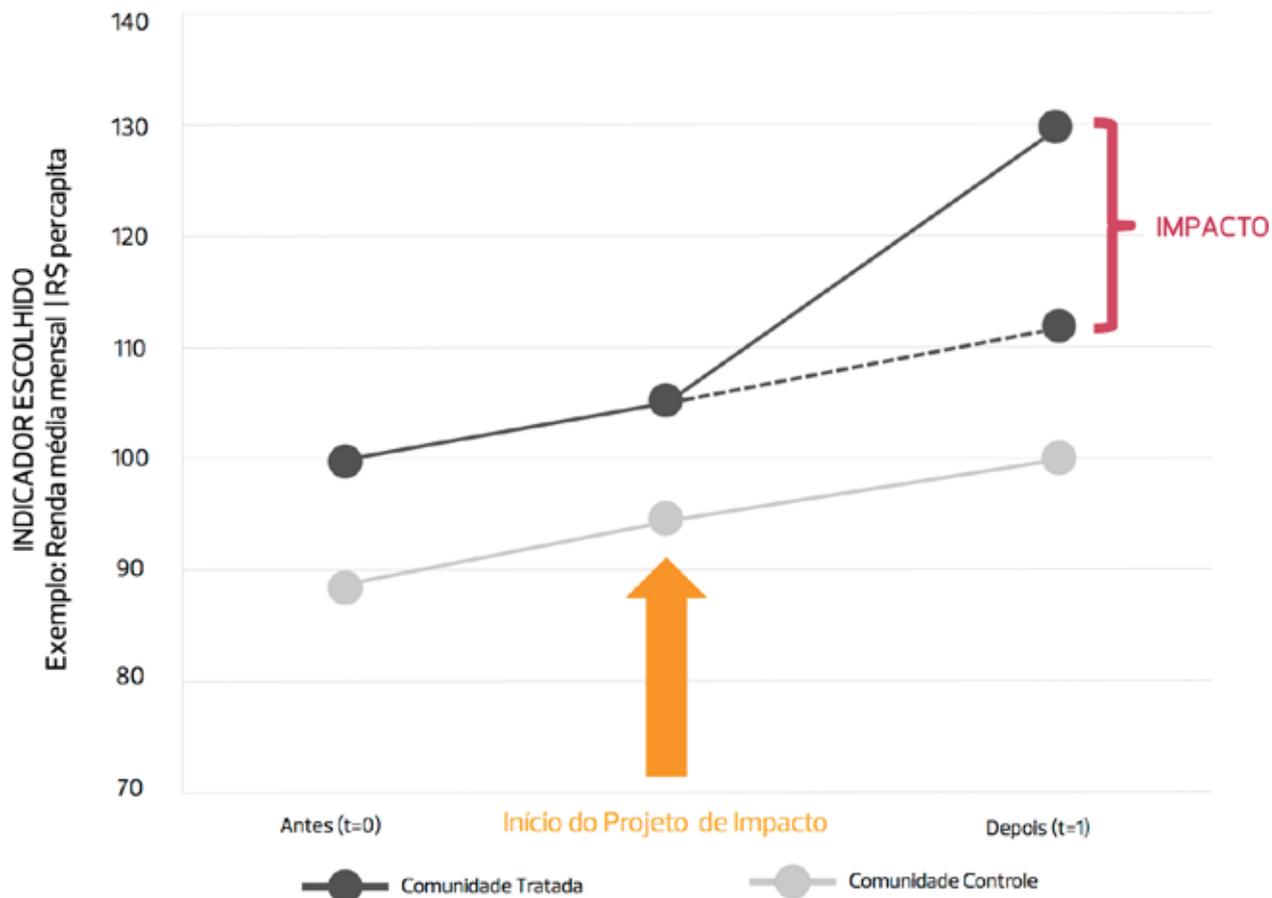


Figura 1: Princípio da Medição por Verificação de Adicionalidade

Nesse sentido, monitorar indicadores não é a mesma coisa que avaliar o seu impacto. Por exemplo, suponha que uma organização invista em atividades diversas para uma comunidade e estabeleça uma meta de que a renda da comunidade deva dobrar em dois anos. Ainda que dois anos após a renda da comunidade tenha dobrado, essa mudança pode ser resultado do crescimento econômico da região e não do projeto realizado. Neste caso, o monitoramento das atividades pode erroneamente apontar que a meta foi alcançada, quando em realidade o desempenho do projeto foi causado por fatores externos.

Idealmente, indicadores devem representar aspectos fortemente alinhados a melhorias concretas na população-alvo. Por exemplo, estudos indicam que o desempenho de alunos do ensino médio e fundamental em testes padronizados têm uma forte relação com o ganho de renda deles no futuro. Logo, o grau com que o investimento de impacto aumenta o desempenho dos alunos nesses testes padronizados tende a ser uma dimensão de alta relevância para a avaliação do impacto.

Seguindo essa abordagem de adicionalidade, o presente guia se baseia em dois pilares essenciais:

- 1. Indicadores com elevada relevância para aferir o impacto do projeto, isto é, resultados positivos esperados na população-alvo.**
- 2. Avaliação com o que poderia ter acontecido a essas populações na ausência do investimento (por exemplo, por meio de comparação com grupos de controle similares que não foram contemplados pelo projeto).**

Dessa forma, o *Guia para Avaliação de Impacto Socioambiental para Utilização em Investimentos e Negócios de Impacto* tem a finalidade de prover um ferramental prático que possibilite a fundos, empresas e governos avaliar o impacto que causam os projetos em que investem, atraindo e alavancando, dessa forma, investidores interessados em impacto socioambiental. Com métricas mais claras e estritamente ligadas aos projetos, que podem ser validadas por terceiras partes independentes, é possível atrair mais recursos para esses projetos e até mesmo estabelecer contratos de metas entre os apoiadores e os executores dos projetos.

ELABORAÇÃO DO PLANO DE MEDIÇÃO

O plano de medição é essencial para planejar, operacionalizar, executar e controlar a medição. Com a finalidade de implementar uma medição robusta e acurada, é importante seguir as seguintes etapas:

ETAPA 1: DELIMITAÇÃO DA POPULAÇÃO-ALVO

O primeiro passo na definição de um plano de medição é a delimitação do setor de atuação do projeto de impacto e especialmente sua população-alvo. Por exemplo, um projeto pode ter como objetivo aumentar o acesso dos cidadãos de baixa renda a medicamentos e formas de tratamento efetivas de doenças crônicas a um menor custo.

A delimitação da população-alvo pode ser feita tanto em termos regionais quanto com base em atributos específicos ligados a potenciais vulnerabilidades (renda, nível educacional, incidência de doenças, dentre outros). Tal definição afetará o tamanho amostral necessário para a realização da medição. Seguindo com nosso exemplo do projeto relativo ao setor de saúde, é possível especificar, por exemplo, que o investimento ocorrerá em postos de saúde e hospitais em bairros selecionados das Zonas Norte e Leste de São Paulo, com foco em famílias com renda mensal abaixo de um determinado limite e com elevada incidência de doenças crônicas. É também interessante ligar os objetivos gerais do projeto aos Objetivos de Desenvolvimento Sustentável (ODSs) das Nações Unidas, que têm sido crescentemente usados para identificar questões-chave para aumentar o bem-estar de populações ao redor do mundo.¹

ETAPA 2: BENCHMARKING

A expansão dos investimentos de impacto depende de um esforço de aprendizado contínuo e consolidação de erros e acertos de projetos similares, bem como do uso de conhecimento prévio de medição gerado por pesqui-

sas e estudos. Dessa forma, a análise de casos de sucesso prévios no setor de impacto (*benchmarking*) é essencial: antes de se definir indicadores, é preciso levantar o que já foi feito ou estudado na área do projeto, sobretudo quais os indicadores foram previamente utilizados em projetos semelhantes.

Fontes de informação incluem: estudos acadêmicos, publicações de organizações internacionais, relatórios de outros investidores de impacto que realizaram aplicações no mesmo setor, dentre outros. Em estudos de impacto, tem sido também cada vez mais comum realizar "meta análises": estudos que tentam agregar os resultados de diversas avaliações prévias sobre um determinado tema.

ETAPA 3: DEFINIÇÃO DA TEORIA DA MUDANÇA

Antes de aprofundar a questão de como medir, primeiramente é necessário definir o que medir. Nessa etapa, o projeto/programa/negócio a ser avaliado deve apresentar sua *teoria da mudança*: uma forma clara e lógica de articular a conexão entre as atividades realizadas e o objetivo socioambiental pretendido.

A **Figura 2** apresenta as cinco etapas de uma teoria da mudança, cujo mapeamento é essencial em um plano de medição: **insumos** e **atividades**, que geram resultados imediatos ou **produtos (outputs)**; que, por sua vez, geram **resultados (outcomes)**, ligados tanto às atividades principais do projeto, quanto a transformações mais amplas para a sociedade. Muito embora grande parte dos projetos socioambientais realize monitoramento de insumos, atividades ou produtos, é crucial para a análise de impacto avaliar mudanças efetivas na população-alvo.

¹ Veja em <http://www.un.org/sustainabledevelopment/sustainable-development-goals/> (accessed on June 30, 2017).



Figura 2: Teoria de Mudança Aplicada a Investimentos e Negócios de Impacto

Fonte: Adaptado de McLaughlin e Jordan; Yoong, P. S., com base em Hehenberger; Harling; Scholten²

Para ilustrar a aplicação da teoria da mudança, vale retornar ao exemplo anterior de uma empresa cuja atividade seja fornecer ferramentas tecnológicas para melhorar os resultados acadêmicos de alunos da rede pública de ensino, utilizando insumos diversos como as próprias ferramentas fornecidas, com o apoio de equipes de suporte para coordenar a sua aplicação. Os produtos dessa atividade seriam o número de acessos à ferramenta na escola ou o número de sessões de interação com alunos.

Observe, entretanto, que nenhum desses resultados indica necessariamente que os alunos melhoraram como resultado da intervenção. É preciso, portanto, analisar quais são os resultados esperados na população-alvo, isso é, melhorias que efetivamente podem causar impacto positivo na vida dos beneficiários ou na sociedade de forma mais ampla. Nesse caso, um resultado ligado à atividade do projeto seria a melhoria do aprendizado dos alunos. Um resultado social mais amplo, por sua vez, seria o aumento esperado de bem-estar e até mesmo ganhos de renda futura em função das novas habilidades adquiridas pelos alunos.

Nesse processo, é importante descrever o **mecanismo causal** que irá determinar as mudanças. Por exemplo, as ferramentas podem incluir jogos e vídeos, ajudando a motivar os alunos para aprendizado e facilitando a compreensão e retenção de conhecimento. Feito esse mapeamento, à luz da Etapa 2, será mais fácil definir indicadores relevantes e claros para o impacto do projeto avaliado. Vale notar que, sob uma ótica de adicionalidade, a teoria de mudança, em si, não avalia o impacto do projeto. O impacto, como discutido na Etapa 5, envolve resultados que ocorreram devido ao projeto, e não a outros fatores externos que podem ter afetado a população-alvo.

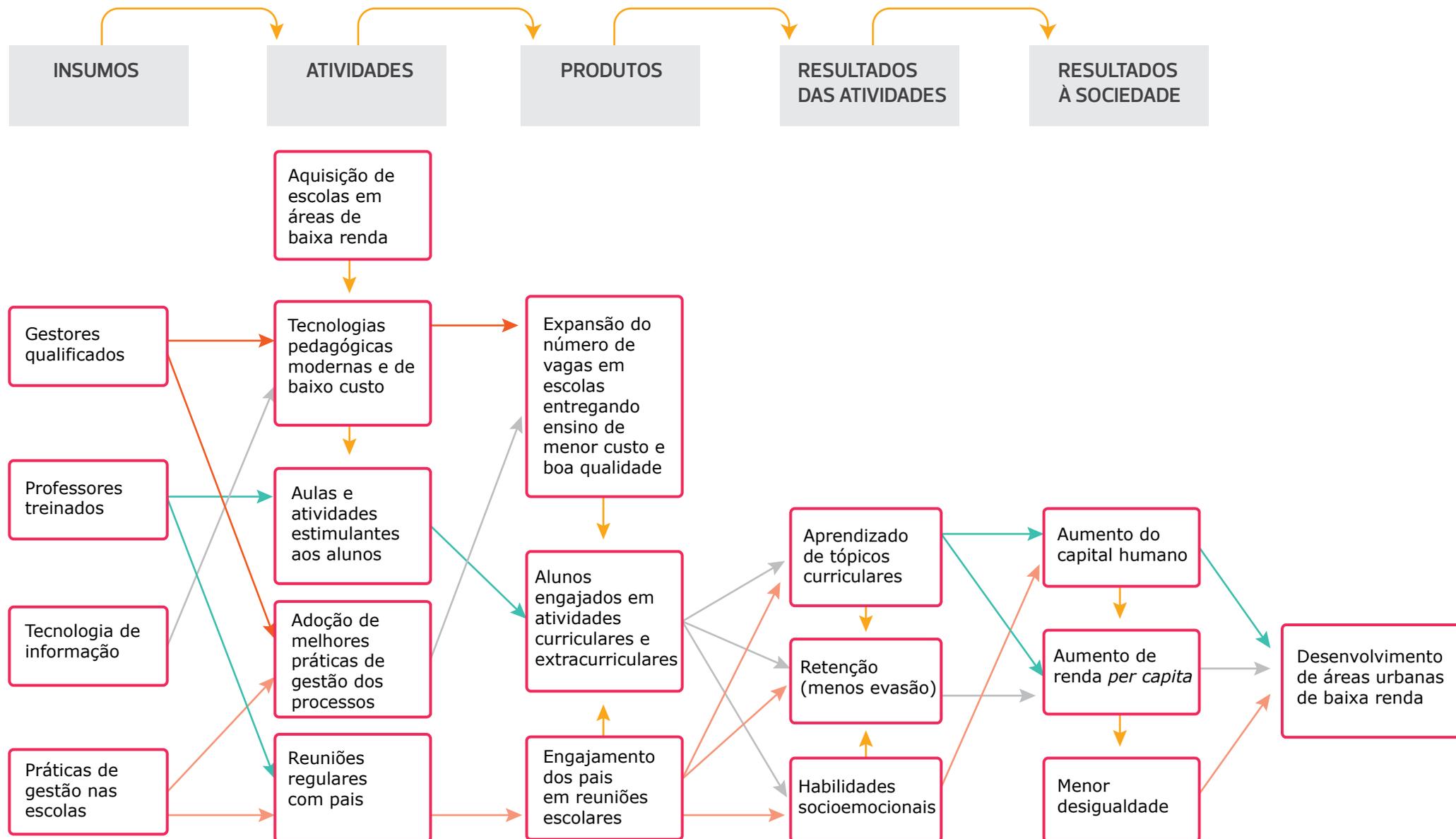
O **Box 1** apresenta um outro exemplo de aplicação de teoria de mudança na área educacional, incluindo um conjunto mais amplo de insumos e intervenções e com uma especificação mais detalhada das possíveis relações de causa e efeito entre os elementos da teoria.

No exemplo a seguir, um investidor de impacto adquiriu um conjunto de escolas privadas de baixa renda e estabeleceu um conjunto de atividades pedagógicas e de gestão para aumentar o aprendizado e a retenção dos alunos.

² MCLAUGHLIN, J. A.; JORDAN, G. B. *Using logic models*. In: *Handbook of Practical Program Evaluation*, 2004, pp. 7-32; HEHENBERGER, L; HARLING, A.M.; SCHOLTEN, P. *A Practical Guide to Measuring and Managing Impact*, 2013, p.9. p.9.

BOX 1.

Exemplo de teoria de mudança para uma rede de ensino médio privada voltada a alunos de baixa renda.



ETAPA 4: DEFINIÇÃO DAS MÉTRICAS (INDICADORES DE IMPACTO)

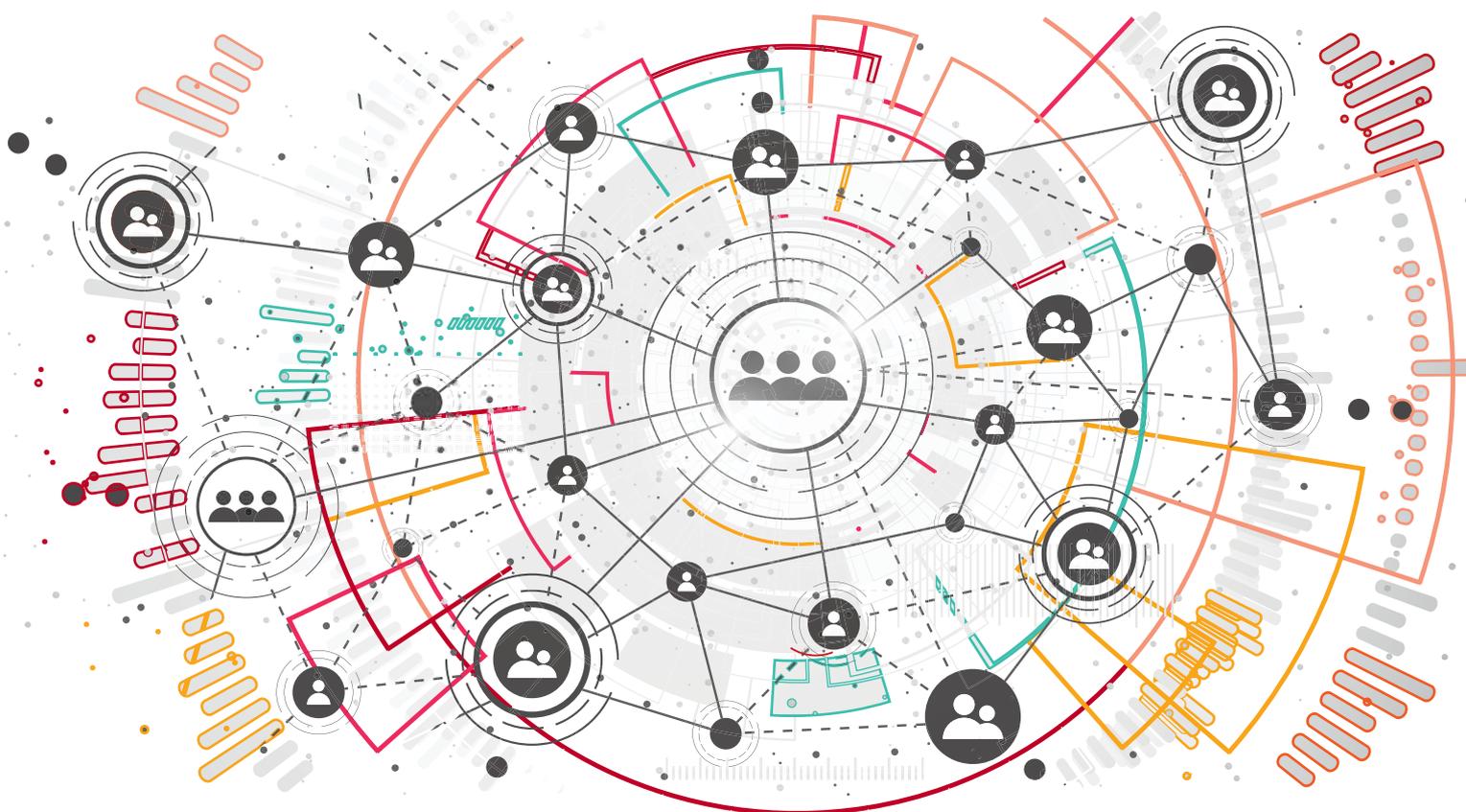
Considerando os resultados das Etapas 2 e 3, devem ser propostos indicadores para o projeto. Como forma de preservar foco e facilidade de compreensão da análise, recomenda-se não mais que três métricas de elevada relevância para avaliar o impacto gerado. Esses indicadores devem ser intimamente ligados às melhorias de curto ou longo prazo identificadas na teoria de mudança (Etapa 3). Métricas ideais apresentam as seguintes características:³

- Têm elevada **relevância** em termos de geração de impacto. Isto é, representam resultados altamente desejáveis para a população-alvo. Se a métrica deriva da teoria da mudança realizada na Etapa 3 e se a teoria corretamente identifica resultados desejáveis na população, então ela deve ser relevante.
- São **passíveis de serem afetadas** pela ação dos executores do projeto. Não podem, por exemplo, ser indicadores muito amplos cujo resultado depende de muitos fatores sem controle direto dos executores. Em geral, métricas ligadas a resultados das atividades do projeto são mais passíveis de ação do que métricas mais ligadas a transformações sociais amplas.

- Devem ser medidas e verificadas com relativa **precisão**. Em geral, indicadores baseados em dados objetivos são preferíveis a indicadores muito subjetivos e com elevado erro de medição. Da mesma forma, os dados não podem ser passíveis de manipulação por gestores, apoiadores e outras partes interessadas em uma boa avaliação do projeto.

- Devem ter **baixo custo** de medição – considerando, inclusive, o custo de obter dados em populações utilizadas como comparação (grupo de controle). O custo da medição tende a se reduzir com o uso de dados secundários, publicamente disponíveis, em comparação com dados primários, coletados especificamente para o projeto.

Como sugestão, ao se considerar diferentes medidas alternativas, pode-se avaliar cada métrica com base nos itens acima e escolher aquelas que mais atendem a maior parte desses critérios. Infelizmente, muitas vezes não é possível atender a todos os critérios, devendo os gestores optar por métricas que melhor se encaixam aos objetivos do projeto. O **Box 2** apresenta um exemplo de análise comparativa de possíveis métricas, identificando visualmente suas vantagens e desvantagens.



³ ROBERTS, J. *Designing Incentives in Organizations*. *Journal of Institutional Economics*, 6 (1): 125–132, 2010.

BOX 2.**Exemplo de análise de adequação de métricas alternativas para um projeto de parque turístico.**

O exemplo a seguir ilustra a escolha de possíveis métricas para um parque em uma área de conservação florestal e com atrações turísticas. O governo está interessado em estabelecer metas contratuais de impacto para guiar as ações de uma empresa privada escolhida para gerir o parque. Há, em particular, uma preocupação em gerar desenvolvimento para as comunidades nos municípios de entorno, além de indicadores de preservação ambiental. Na análise abaixo, as métricas ligadas à renda são relevantes, precisas e de baixo custo, porém a métrica de ganho de renda ligada a atividades de turismo é mais passível de ação, pois a empresa gestora pode contratar

pessoas da região para trabalhar no parque ou estimular empreendedores a ofertar serviços complementares (como hotéis e restaurantes). O indicador de satisfação dos usuários, apesar de relevante e passível de ação, tem um alto custo e é menos preciso – além de se basear em percepções subjetivas, traz risco de manipulação (por exemplo, gestores podem influenciar que a coleta seja feita em momento de menor lotação do parque, reduzindo a incidência de queixas). O indicador com o percentual de área seguindo regramentos ambientais, se informado pela empresa gestora, pode exigir contratação de auditoria técnica especializada, o que tende a aumentar os custos.

	RELEVANTE?	PASSÍVEL DE AÇÃO?	PRECISA?	DE BAIXO CUSTO?
1. Ganho de renda da população local	✓ Desenvolvimento local é um dos objetivos do governo com o projeto.	✗ A economia local pode ser afetada por outros fatores além das atividades do parque.	✓ Há estatísticas disponíveis e confiáveis sobre renda local.	✓ As estatísticas são amplamente divulgadas pelo governo e de fácil acesso.
2. Ganho de renda local com atividades ligadas ao turismo	✓ Desenvolvimento local é um dos objetivos do governo com o projeto.	✓ O parque deve influenciar a movimentação de hotéis e restaurantes, por exemplo.	✓ Há estatísticas disponíveis e confiáveis sobre renda local.	✓ As estatísticas são amplamente divulgadas, de fácil acesso e desagregadas por setor de atividade.
3. Satisfação dos usuários do parque (a partir de pesquisa de opinião)	✓ Uma maior satisfação pode indicar a qualidade dos serviços aos visitantes.	✓ Esforço para prover melhores serviços tende a afetar positivamente a satisfação.	✗ Indicadores de satisfação são subjetivos e há risco de manipulação.	✗ Elevado custo para realização de pesquisas e dificuldade de obter dados de comparação.
4. Percentual da área do parque seguindo regramentos de conservação	✓ Estando em área de conservação, ações de preservação ambiental são centrais.	✓ Gestores do parque podem controlar diretamente o estado das áreas de conservação.	✓ Há protocolos claros indicando aderência a padrões de conservação.	⚠ Gestores realizam esse monitoramento, mas pode ser necessária validação técnica externa.

ETAPA 5: DEFINIÇÃO DO MÉTODO DE VERIFICAÇÃO DE ADICIONALIDADE E DO NÍVEL DE MEDIÇÃO

Após a definição das métricas, é importante definir como será feita a verificação da adicionalidade, isso é, considerar o que provavelmente teria acontecido com a população-alvo caso o projeto não tivesse sido executado. Nessa etapa, são definidos o grupo dos **tratados** (indivíduos e comunidades beneficiadas pelo projeto) e o **grupo de controle** (indivíduos e comunidades que não receberam as intervenções).

A forma de verificação de adicionalidade varia de acordo com o **nível de medição** definido para o projeto, conforme detalhado na seção seguinte. Níveis mais elevados de medição trazem mais confiança de que mudanças na população-alvo foram realmente causadas pelo projeto, porém tendem a ser mais complexos e custosos. Nessa etapa, especialmente caso sejam adotados níveis de medição mais elevados, recomenda-se o apoio de especialistas para ajudar na identificação das técnicas mais apropriadas e suas potenciais limitações.

Uma vez definida a forma de verificação de adicionalidade, e com base nos indicadores escolhidos, é importante definir com precisão como será calculado o impacto. No exemplo da **Figura 1**, o impacto será medido como a diferença de variação de renda entre a comunidade tratada e a de controle, considerando os dois períodos definidos.

ETAPA 6: PLANO AMOSTRAL

A definição do tamanho amostral – quantas pessoas ou grupos participarão da avaliação do projeto – é outra crucial etapa do plano de medição. Essa etapa é particularmente importante no caso de níveis de medição mais elevados, empregando técnicas que se propõem a identificar com mais precisão o efeito do projeto sobre a população-alvo. Com um adequado tamanho amostral, é possível verificar com maior precisão estatística a diferença entre o resultado medido antes e depois do projeto, bem como a diferença entre os grupos tratados e de controle.

Essencialmente, um maior tamanho da amostra reduz o

risco de as análises não encontrarem um efeito positivo do projeto, supondo que realmente ele exista. Em geral, o tamanho da amostra depende do tamanho do efeito esperado: tudo o mais constante, quanto menor o impacto esperado do projeto, maior o tamanho necessário da amostra para que se possa medi-lo. Sempre que possível, é também importante procurar ter grupos tratado e de controle com tamanho amostral relativamente similar. No entanto, os custos de realizar o tratamento e de coletar os dados relevantes devem ser levados em consideração para otimizar a amostra.⁴

ETAPA 7: EXECUÇÃO DA MEDIÇÃO

Sempre que possível, de forma a reduzir os custos com a coleta de dados, recomenda-se o uso de dados secundários (bases de dados confiáveis e publicamente disponíveis). Havendo necessidade de coletar dados originais (primários), é imprescindível contar com o apoio de uma equipe treinada e especializada em pesquisa de campo.

Nesse caso, para fins de verificação de adicionalidade, é importante lembrar que os dados devem ser coletados não apenas nas populações afetadas pelo projeto, mas também no grupo de controle. É também importante que, no plano de medição, o executor do projeto evidencie quais serão as métricas, quem irá executar a medição e qual será o orçamento de todo processo.

ETAPA 8: CRONOGRAMA DE MEDIÇÃO

Nessa etapa, é importante definir o período de coleta de dados, indicando, se for o caso, as medições antes e depois do início das intervenções (como indicado na **Figura 1**). O executor do projeto e o investidor devem ter consciência de que, na maioria dos casos, a verificação de impacto ocorrerá após o período de, no mínimo, um ou dois anos desde o início do projeto. Algumas áreas de investimento requerem um horizonte ainda maior para possibilitar a real verificação do impacto. Por exemplo, no setor de educação, em que o aprendizado dos alunos requer um esforço longo e persistente, o horizonte de tempo tende a ser de dois ou três anos, pelo menos, e em diversos casos até maior.

⁴ Uma discussão técnica sobre como definir o tamanho da amostra pode ser vista no texto de DUFLO, E.; GLENNERSTER, R.; KREMER, M. *Using Randomization in Development Economics Research: a Toolkit*. Mimeo, 2013. Disponível em <http://www.nber.org/papers/t0333>. Uma ferramenta prática para cálculo de tamanho amostral pode ser obtida em <https://sites.google.com/site/optimaldesignsoftware/home> (acesso em 9/1/2018). Sobre técnicas e cuidados de avaliação, consulte também ANGRIST, J. e PISCHKE, J.-S. *Mostly Harmless Econometrics*. Princeton: Princeton University Press, 2009; e MENEZES FILHO, N. A. (editor) *Avaliação Econômica de Projetos Sociais*. 1. Ed. São Paulo: Dinâmica Gráfica e Editora, 2012.

NÍVEIS DE MEDIÇÃO

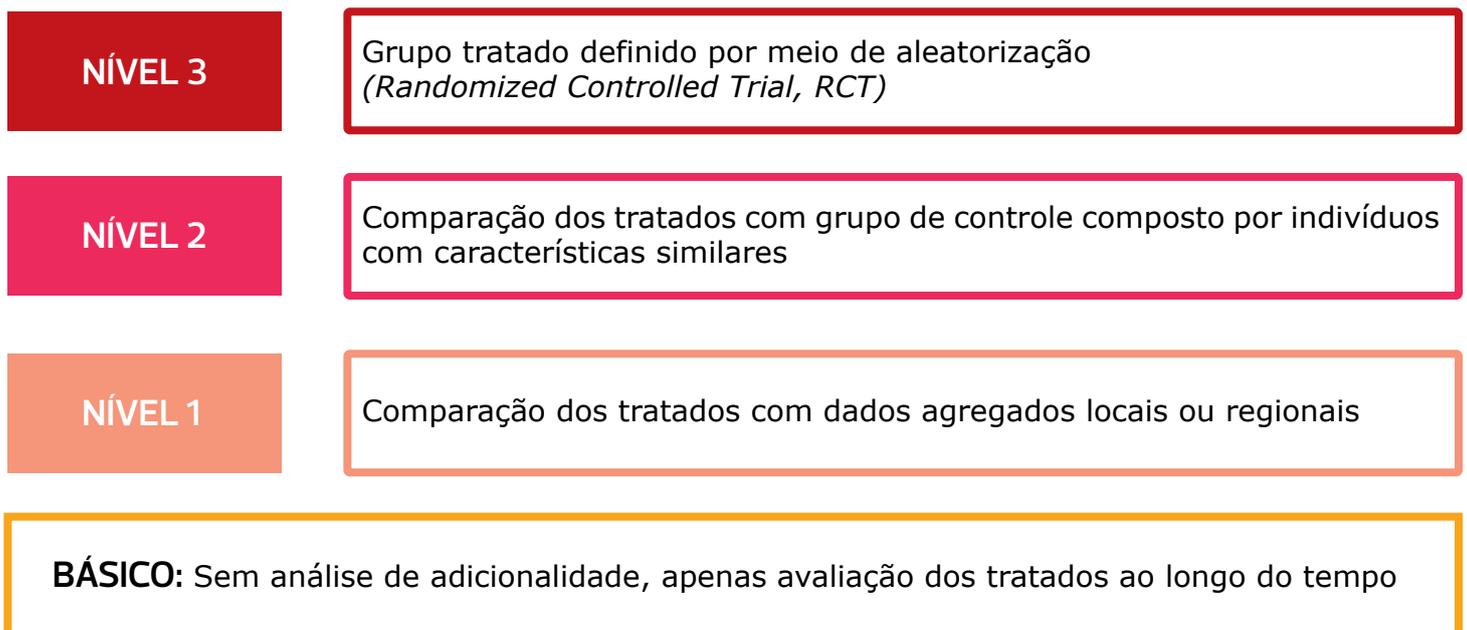
Os níveis de medição, essencialmente, propõem formas alternativas de corrigir o efeito de fatores externos decorrentes de diferenças pré-existentes entre os grupos tratados e de controle. Por exemplo, imagine que o investidor de impacto queira avaliar o efeito de um novo método pedagógico e ofereça esse método a diversas escolas. É possível que as escolas que aceitem aderir ao programa sejam justamente aquelas com gestão mais motivada para melhorias no ensino. Se essa variável afetando a escolha ("gestão motivada") não for observada pelo avaliador, pode-se concluir erroneamente que possíveis melhorias nas escolas que voluntariamente aderiram ao programa resultaram do programa em si, ainda que a causa real tenha sido a propensão natural das escolas escolhidas a melhorar o seu ensino.

Para minimizar esse erro, especialistas em avaliação com verificação de adicionalidade enfatizam que, idealmente, os grupos e indivíduos tratados devem ser escolhidos de forma aleatória, por sorteio. No entanto, sabe-se que o sorteio de indivíduos ou grupos é particularmente complexo e em muitos casos até mesmo inexecuível. A decisão pelo investimento em determinados projetos, seja

essa tomada por fundos, indivíduos, empresas ou governos, na maior parte dos casos, segue pressupostos e planos previamente estabelecidos. Assim, em muitos casos, os investidores e executores de projetos de impacto acabam tendo que recorrer a outros métodos que tentam corrigir diferenças prévias entre os grupos tratado e controle.

Qualquer que seja a forma de verificação escolhida, é importante dar total transparência sobre o método adotado e suas possíveis limitações. Especificamente, abaixo são propostos **níveis (tiers) de medição**, de acordo com o grau de robustez pretendido para a avaliação de impacto, em especial no que tange ao rigor sobre a aferição do efeito *causal* do projeto. Na **Figura 3**, partindo-se da base ao topo, inicia-se com um tipo de medição denominado de *básico*, que não envolve análise de adicionalidade, como forma de expor como cada nível permite maior rigor na estimativa do efeito do projeto. Uma vez que negócios e investimentos de impacto normalmente utilizam mais de uma métrica, é possível que os próprios indicadores sejam avaliados de acordo com níveis de medição distintos.

Figura 3: Níveis de Medição para Verificação de Adicionalidade



BÁSICO – MENSURAÇÃO APENAS DOS RESULTADOS DO GRUPO TRATADO

Essa abordagem *não* envolve cômputo de adicionalidade e *não* deve ser considerada como um nível de medição, uma vez que não inclui análise de cenário contrafactual (o que teria acontecido com a comunidade-alvo sem a intervenção). Entretanto, para muitas métricas, é a forma mais factível de ser executada, uma vez que é mais fácil coletar dados dos indivíduos ou grupos participando do projeto. Nesse caso, observa-se simplesmente como determinados indicadores variaram no início e ao longo das intervenções.

É também uma abordagem bastante usada para acompanhar indicadores de desempenho do próprio projeto. Contudo, ao observar variações positivas nesses indicadores, os gestores devem ter cuidado de não necessariamente inferir que essas mudanças foram causadas pelas intervenções realizadas. A própria população-alvo pode já ter uma tendência de melhoria mesmo antes do início das intervenções e podem ter ocorrido mudanças externas favorecendo os resultados do projeto.

NÍVEL 1 – MENSURAÇÃO COMPARANDO O GRUPO TRATADO COM DADOS AGREGADOS LOCAIS OU REGIONAIS

Nesse nível, empregam-se dados agregados como forma de comparação dos resultados do projeto ao que poderia ter ocorrido sem a sua implantação. Antes da medição, há uma definição deliberada da localidade onde o projeto será realizado e de quem são os indivíduos beneficiados. Além disso, para fins comparativos, são utilizados dados agregados já disponíveis para uma determinada região onde é feito o investimento. Devem ser buscados dados que permitam avaliar o que ocorreu com o projeto comparativamente ao resultado agregado antes e depois da intervenção (**Box 3**).

BOX 3.

Exemplo de verificação de adicionalidade no Nível 1

Uma empresa agroindustrial decide realizar compras de produtos agrícolas de pequenos produtores familiares em certos municípios de baixa renda, ao mesmo tempo apoiando esses produtores com assistência técnica e gerencial. A definição das comunidades beneficiadas é feita de forma deliberada pelo executor do projeto.

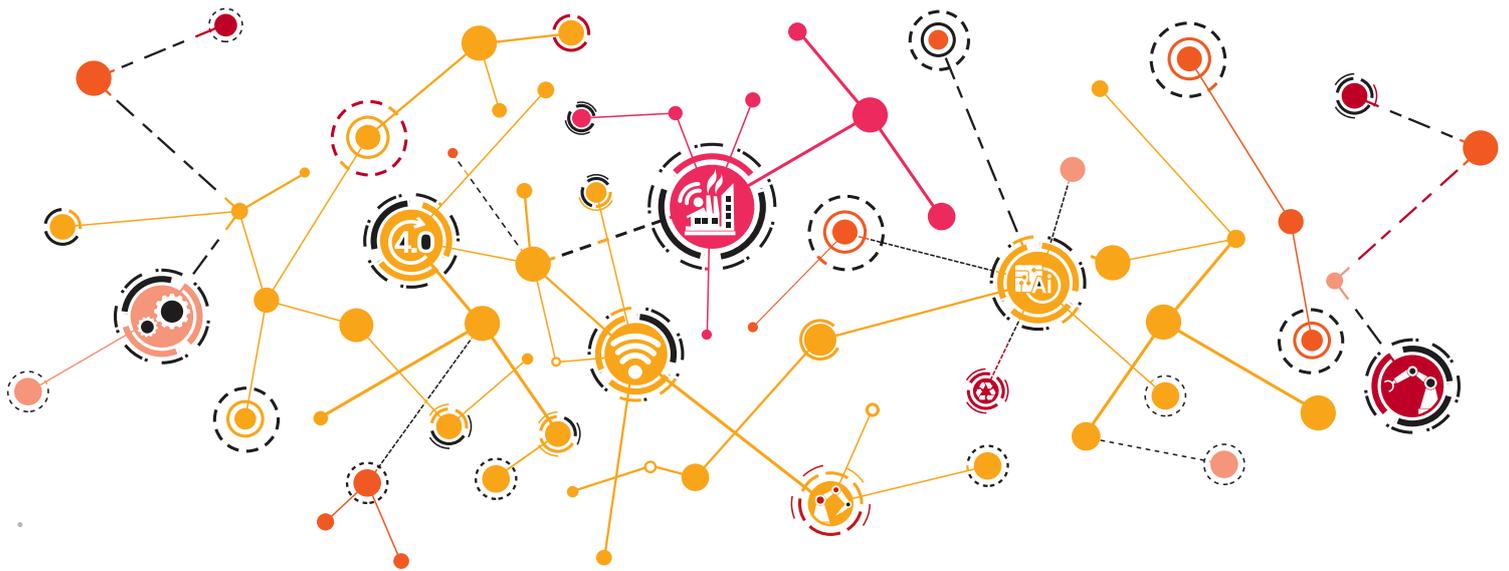
Utilizando a renda dos produtores como uma métrica de impacto, a avaliação é efetuada comparando-se a renda das famílias apoiadas frente à renda de famílias em áreas rurais do município ou microrregião onde foi feito o investimento, antes e depois da intervenção. Será verificada adicionalidade no Nível 1 caso a variação de renda das famílias apoiadas supere a variação agregada das famílias na região no mesmo período.

NÍVEL 2 – MENSURAÇÃO UTILIZANDO UM GRUPO DE CONTROLE COM CARACTERÍSTICAS SEMELHANTES ÀS DO GRUPO TRATADO

Nesse nível, embora ainda não seja implementada aleatorização, há uma tentativa de criar um grupo de controle formado por indivíduos ou comunidades comparáveis aos tratados. Por isso, as técnicas de verificação de adicionalidade ao Nível 2 são normalmente chamadas de **quase-experimentais**. Da mesma forma que no Nível 1, há uma definição deliberada da localidade onde o projeto será realizado e de quem são os indivíduos beneficiados. E, no processo de avaliação, empregam-se técnicas para construir um grupo de controle formado por indivíduos que sejam parecidos com os indivíduos tratados. Ou seja: em vez de utilizar dados agregados de um determinado local, há uma tentativa de acompanhar grupos ou indivíduos que não recebem o investimento, porém com características similares às dos tratados.

A forma mais usual de escolher indivíduos com características semelhantes entre os grupos de tratados e de controle envolve a chamada técnica de *pareamento*.

Nela, com base em características observáveis relevantes para a seleção de participantes na intervenção (como idade, sexo, renda e assim por diante), procuram-se uma ou mais unidades no grupo do controle que sejam as mais parecidas quanto possível a cada unidade no grupo de tratamento. Em casos em que exista competição para participar no projeto e um determinado critério de seleção, uma possibilidade é considerar a *descontinuidade* gerada pela própria seleção. Isso é, podem ser comparados os participantes que quase não foram aceitos (tratados) àqueles que quase foram aceitos, mas que, eventualmente, ficaram abaixo do limite de corte (controle). Em situações em que se observa o impacto em um caso específico (uma única organização ou comunidade, por exemplo), há ainda a possibilidade de criar um "controle sintético" a partir de uma combinação de outros casos que mais se aproximem do caso tratado.⁵ O **Box 4** apresenta exemplos de verificação de adicionalidade no Nível 2.



No Nível 2, o princípio geral é garantir que os grupos de tratamento e controle sejam realmente comparáveis, tanto em termos de seus atributos quanto em termos da própria evolução passada do seu desempenho. Por exemplo, na medição exemplificada na **Figura 1**, um importante aspecto é que os grupos tratado e de controle exibam uma tendência similar de evolução da métrica de impacto antes da intervenção. Tendências não alinhadas poderiam sugerir a existência de fatores distintos (e não

controlados) influenciando a evolução dos grupos. No caso de análises com base em descontinuidades, como no **exemplo 4.2** do **Box 4**, o avaliador pode testar se os grupos diferem em termos de outras variáveis além da variável que marca a descontinuidade. Por exemplo, microempreendedores logo acima ou abaixo do critério de corte não deveriam diferir em termos de volume de vendas, renda e outras variáveis relevantes que poderiam afetar os resultados observados.

BOX 4.

Exemplos de verificação de adicionalidade no Nível 2

Exemplo 4.1: Em um projeto de impacto cujo indicador é a renda de famílias, a avaliação é efetuada comparando-se a renda das famílias tratadas com dados sobre a renda de famílias no grupo de controle. A definição das comunidades beneficiadas pelo projeto é feita de forma deliberada pelo executor. Neste nível de medição, as famílias de controle são então selecionadas com base em características observáveis mais parecidas quanto possível àquelas das tratadas. Por exemplo, no grupo de controle podem ser selecionadas famílias com níveis iniciais de renda, escolaridade e número de filhos similares aos níveis observados nas famílias tratadas.

Exemplo 4.2: Uma organização pretende implementar um programa de microcrédito, convidando microempreendedores a participarem do programa, pré-avaliados com base em indicadores de risco de crédito (renda, histórico em empréstimos anteriores, etc.). Havendo vagas limitadas, apenas os mais bem avaliados serão selecionados, a partir de um critério de corte estabelecendo um nível mínimo de risco de crédito. No limiar desse critério de corte, pode-se supor que os indivíduos são parecidos, dado que apresentam avaliação similar de risco de crédito. Dessa forma, pode-se comparar o grupo de microempreendedores selecionados logo acima da linha de corte com aqueles que não foram selecionados, mas que ficaram logo abaixo da linha.

⁵ Um exemplo de forma de implementação de técnica de pareamento pode ser vista em: ABADIE, A.; DRUKKER, D.; HERR, J. L.; IMBENS, G. W. *Implementing Matching Estimators for Average Treatment Effects Stata*. *The Stata Journal*, 4(3): 290–311, 2004. Para uma discussão mais geral de vários métodos alternativos, consulte ANGRIST, J. e PISCHKE, J.-S., op. cit. A técnica de controle sintético é discutida em ABADIE, A.; DIAMOND, A.; HANMUELLER, J. *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program*. *Journal of the American Statistical Association*, 105(490): 493–505.

NÍVEL 3 – MENSURAÇÃO COM ALEATORIZAÇÃO (RANDOMIZED CONTROLLED TRIAL, RCT)

As técnicas do Nível 3, também chamadas de **experimentais**, permitem máxima confiabilidade da estimativa de avaliação de impacto. Neste nível, a decisão de quais indivíduos ou comunidades serão sujeitos à intervenção é feita de forma totalmente aleatória, logo reduzindo possíveis vieses de seleção dos grupos-alvo do projeto.

Em casos em que não é possível ou desejável excluir determinados grupos, pode-se implementar um desenho alternativo em que um grupo é encorajado a adotar o que é proposto pela intervenção, sendo que a atribuição do encorajamento é definida de forma aleatória entre os indivíduos ou comunidades-alvo (**Box 5**). A escolha aleatória, se bem-feita, garante que não existam diferenças entre as características, observáveis ou não, dos grupos de tratados e de controle. Por isso, medições do Nível 3 em geral não necessitam de medição em um período inicial, anterior ao projeto, como indicado na **Figura 1**. Seguindo-se procedimentos adequados de aleatorização, pode-se avaliar o desempenho comparativo dos grupos tratados e de controle em um momento após o início do projeto. No Nível 3, é particularmente importante atentar para cálculos de tamanho mínimo da amostra, conforme discutido na Etapa 6 da seção anterior.

É importante não confundir amostragem aleatória com definição aleatória de tratamento. No Nível 3, a aleatorização se dá na escolha do grupo que deve receber ou não a intervenção. Por exemplo, suponha que determinado projeto de educação tenha beneficiado 30 escolas, sendo o processo de escolha definido de forma não aleatória. Então, o Nível 3 não será atingido mesmo se forem selecionadas, aleatoriamente, outras 30 escolas como grupo de controle. Isso porque a seleção de quem recebeu a intervenção não foi feita de forma aleatória, podendo, portanto, ser afetada por fatores não observáveis pelos avaliadores no momento de calcular o impacto da intervenção.

COMPARANDO OS NÍVEIS DE MEDIÇÃO

O **Quadro 1** apresenta um comparativo das vantagens e limitações de cada nível de medição discutidos anteriormente. Em linhas gerais, movendo-se em direção ao Nível 3, é possível aumentar a confiabilidade das medições de impacto, no sentido de garantir que o projeto de fato

BOX 5.

Exemplos de verificação de adicionalidade no Nível 3

Exemplo 5.1: Uma organização oferece um programa limitado de microcrédito em um contexto em que não há recursos para apoiar todos da comunidade. A organização pode convidar interessados e definir quem vai receber o microcrédito de forma aleatória. Alternativamente, pode emprestar para aqueles com melhor risco de crédito e então aleatorizar dentro do conjunto daqueles que quase foram selecionados, isto é, logo abaixo da linha de corte de risco de crédito. Considerando esse conjunto, pode-se definir de forma aleatória um grupo de indivíduos tratados que receberão o crédito, sendo então o grupo de controle definido por aqueles não escolhidos por esse procedimento de aleatorização.

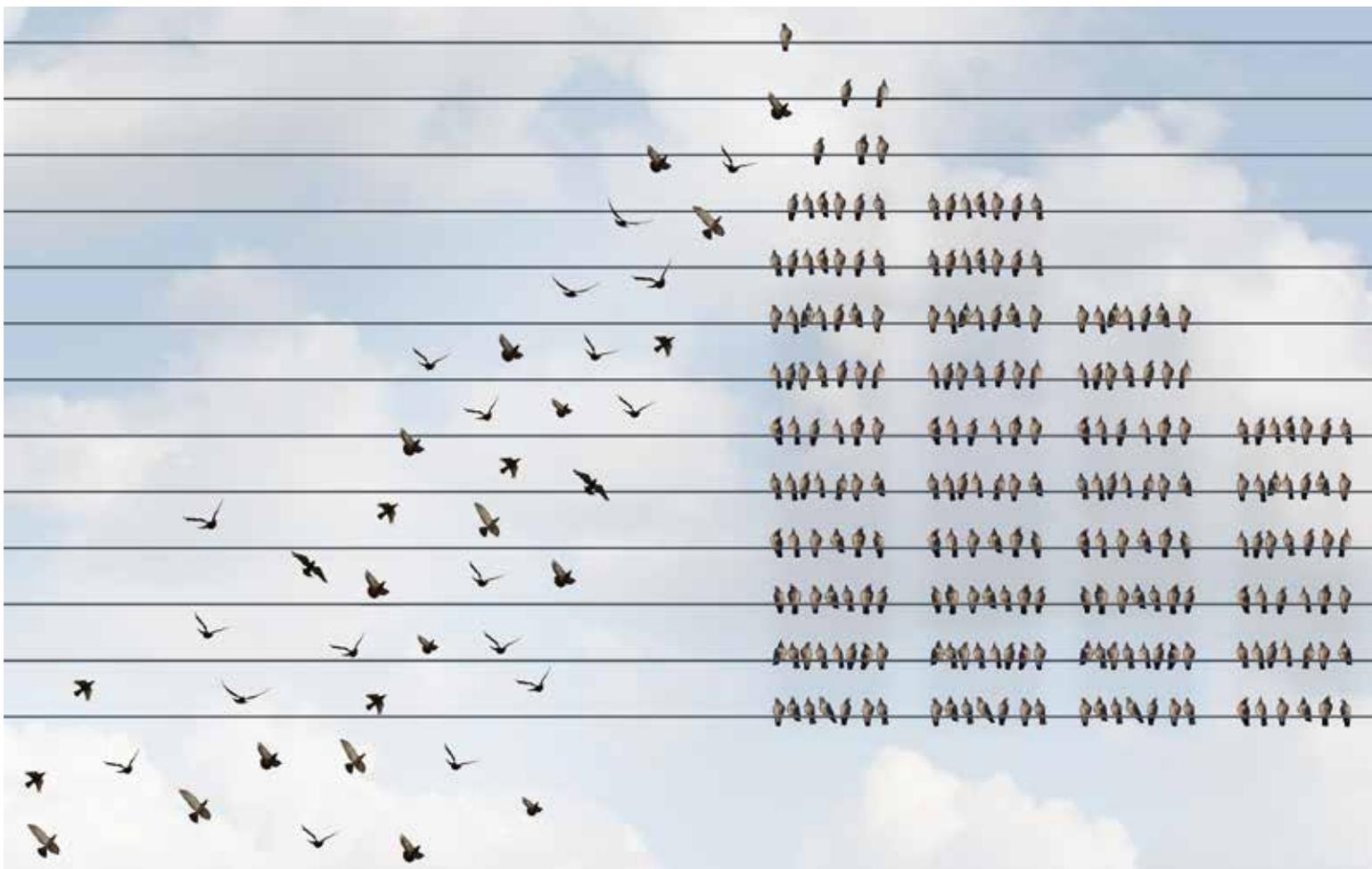
Exemplo 5.2: Um amplo grupo de alunos de uma rede pública recebe acesso a uma plataforma de aprendizado por internet. Ainda que todos os alunos tenham acesso à plataforma, pode-se sortear um grupo de alunos que receberá mensagens estimulando o seu uso. Considerando-se que uma fração desses alunos que recebeu as mensagens usará de fato a ferramenta (ou seja, somente estes serão efetivamente tratados), é possível comparar o desempenho desses alunos ao desempenho do grupo de controle de alunos que não recebeu as mensagens. Essa análise, entretanto, requer ajustes estatísticos, uma vez que se quer avaliar quem efetivamente usou a plataforma, e não necessariamente quem recebeu ou não recebeu o estímulo.⁶

causou as mudanças detectadas na população-alvo. Porém, ao mesmo tempo, aumenta-se a complexidade das análises e do desenho da pesquisa, além de exigir potencialmente mais recursos técnicos e financeiros para as análises.

⁶ Nesse caso, pode-se medir o impacto da intervenção a partir do estimador de efeito médio local do tratamento (*LATE – Local Average Treatment Effect*). Ver ANGRIST, J. e PISCHKE, J.-S., op. cit.

QUADRO 1: Vantagens e limitações dos níveis de medição

	VANTAGENS	LIMITAÇÕES
Básico	<ul style="list-style-type: none"> • Dados mais facilmente disponíveis (coletados do próprio projeto e da sua população-alvo). • Maior simplicidade nas análises (comparação apenas dentro do grupo tratado, antes e depois). 	<ul style="list-style-type: none"> • Não envolve verificação de adicionalidade, pois não há análise de cenário contrafactual (o que teria acontecido com a população-alvo sem o projeto).
Nível 1	<ul style="list-style-type: none"> • Em muitos casos, gestores têm à disposição apenas dados agregados (por exemplo, renda municipal ao invés da renda das famílias do município). • Comparando a outros métodos de verificação de adicionalidade, a análise é mais simples (resultados do projeto versus resultados agregados). 	<ul style="list-style-type: none"> • As características da população agregada podem ser muito distintas da população-alvo, ainda que dentro do mesmo recorte geográfico. • Não permite estimação de efeito causal, dado que não envolve técnicas estatísticas para controle de potenciais diferenças entre grupos.
Nível 2	<ul style="list-style-type: none"> • Em muitos casos, a população-alvo tratada já foi definida, sendo inviável realizar aleatorização. • Ao se construir grupos de controle com características similares aos tratados, evita-se viés gerado por diferenças substanciais entre os grupos com base em traços observáveis. • As diferentes técnicas do Nível 2 permitem realizar inferência estatística sobre o impacto observado no grupo tratado, frente ao grupo de controle. 	<ul style="list-style-type: none"> • As técnicas do Nível 2 só permitem avaliar o efeito causal do projeto se não houver efeito relevante de outros fatores não observados e medidos, o que é difícil de verificar na prática. • É preciso obter dados desagregados de características não apenas daqueles afetados pelo projeto, mas também daqueles no grupo de controle. • Deve haver casos no grupo de controle que sejam próximos, em suas características, aos tratados. Caso contrário, não é possível realizar a comparação.
Nível 3	<ul style="list-style-type: none"> • Ao se realizar a aleatorização, é possível garantir com mais segurança que o efeito medido foi realmente causado pelo projeto. • Em geral, a aleatorização não requer medição antes do início do projeto; os indivíduos e grupos podem ser comparados após a realização das intervenções. 	<ul style="list-style-type: none"> • A técnica exige atenção particular para o tamanho da amostra. Torna-se inviável em contextos com poucos casos recebendo a intervenção. • Desenhos experimentais no Nível 3 são muito sujeitos a problemas de efeitos cruzados, atrito e outros discutidos na seção seguinte do Guia. • Podem surgir dilemas éticos em casos em que não se quer excluir indivíduos de uma determinada intervenção.



CUIDADOS DURANTE A MEDIÇÃO

Algumas precauções são extremamente importantes para a realização satisfatória de uma medição de impacto. Abaixo são descritos de forma pontual alguns aspectos críticos que o executor do projeto deve considerar na realização do projeto de medição.⁷

EFEITOS CRUZADOS

Um dos principais pontos a ser considerado é que deve existir o mínimo possível de **efeitos cruzados** (*spillovers*) entre os indivíduos de grupos diferentes. Por exemplo, considere que se pretenda avaliar o impacto de um projeto empresarial para o aumento de renda em uma determinada localidade. Para tanto, previamente ao projeto, realiza-se um levantamento de dados nessa comunidade e na comunidade vizinha, que não foi considerada para o projeto. O efeito cruzado pode ocorrer se, por exemplo, a comunidade tratada, ao receber renda advinda do investimento, puder dispendar parte dos seus recursos comprando produtos ou transferindo renda à outra

comunidade. Se isto ocorrer, a adicionalidade de impacto do programa seria subestimada.

Esses efeitos cruzados são problemáticos para avaliações de impacto, pois eles podem distorcer o resultado pela contaminação do grupo de controle com a intervenção feita no grupo tratado. Para contornar esse problema, é importante garantir que não haja comunicação, troca de informação ou troca física de bens e serviços entre os indivíduos. Uma das formas de conseguir isso é garantir que os grupos tratados e de controle estejam geograficamente distantes ou isolados, com menor risco de efeitos cruzados.

⁷ Essa seção é fortemente baseada em DUFLO et al., op. cit.

ATRITO

Como a medição de impacto ocorre em dois momentos distintos no tempo, um antes do projeto começar e outro depois de certo tempo de vigência, é de se esperar que ocorra um “descasamento” entre os indivíduos avaliados no momento inicial e os avaliados no momento final, tanto no grupo de controle quanto no grupo tratado. Por exemplo: pode ocorrer que alguns indivíduos observados no momento inicial não residam mais na localidade para avaliação final. A essa falha na coleta de resultados para alguns indivíduos que tenham se evadido, ou não se façam mais presentes no momento final, dá-se o nome de **atrito**. Vale destacar que, apesar da importância da aleatorização para a presente metodologia, ela não garante a inexistência de atrito não aleatório ao fim. Ademais, o problema do atrito reforça a importância de se definir, anteriormente à execução do projeto, uma amostra suficientemente grande para a realização das medições.

Como exemplo prático desse problema, considere um

projeto de impacto para o aumento de renda em uma determinada localidade. Ainda que inicialmente os grupos tenham sido escolhidos de forma aleatória, ao fim do projeto, pode-se deparar com atrito não aleatório decorrente da evasão de indivíduos que não tenham recebido o aumento de renda, frente a indivíduos que tenham recebido e continuem na localidade. Nesse caso, o atrito gera um viés na medição de impacto, pois haverá uma perda não aleatória de dados dos indivíduos do grupo de controle (pela evasão de indivíduos de um determinado perfil) frente ao grupo tratado.

É sempre importante expor as taxas de atrito presentes tanto no grupo tratado quanto no de controle de forma transparente. Ou seja, ao final da medição, é preciso evidenciar quantos indivíduos medidos no primeiro momento não foram encontrados para a medição no segundo momento, tanto no grupo de controle quanto no grupo tratado.

EFEITO HAWTHORNE E EFEITO JOHN HENRY

Por fim, outra das principais limitações que podem ocorrer em uma medição de impacto é a mudança de comportamento dos indivíduos tanto do grupo tratado, quanto do grupo de controle.⁸ O chamado **efeito Hawthorne** pode ocorrer pela percepção dos indivíduos tratados de que estão sob algum tipo de intervenção e, assim, há uma mudança de comportamento em resposta.

Em projetos de educação, por exemplo, pode haver um esforço superior dos professores e dos alunos pela conscientização destes indivíduos de que seu grupo está sendo beneficiado pela intervenção. Já o efeito comportamental no grupo de controle, chamado de **efeito John Henry**, pode ocorrer quando professores do grupo controle sentem-se desafiados e começam a competir com os do grupo tratado para mostrar que também teriam condições de se beneficiar do programa. O problema pode ocorrer também em outra direção: por não receberem tratamento, os professores se desmotivam, gerando uma superestimação do real impacto do projeto. Nos dois casos, o efeito do programa pode ser distorcido.

Ainda que respostas comportamentais sejam sempre

complexas, há formas de desenho de medição que possibilitam minimizar os efeitos Hawthorne e John Henry. Por exemplo, pode-se utilizar três grupos: tratado, controle e *placebo*, neste último se estabelece um grupo que será observado ou que manterá relações com o executor do projeto social sem, contudo, ser efetivamente beneficiado pelo projeto. Considere, por exemplo, um projeto no qual uma grande empresa deseja melhorar a renda de determinadas comunidades por meio da compra de insumos locais. Assim, a empresa não deve apenas considerar uma comunidade de atuais supridores (tratada) e uma comunidade comparável de não supridores (controle). A empresa pode também monitorar uma comunidade placebo de não supridores com as quais tem alguma forma de relação (por exemplo, a empresa pode enviar gestores para conversar com essas comunidades, mas sem estabelecer contratos de suprimento). O impacto medido no grupo tratado deve ser superior ao verificado nos grupos de controle e placebo. Um outro procedimento é que o impacto seja aferido por uma organização independente, para garantir uma ligação menos direta entre a intervenção em si e o processo de medição de impacto.

⁸ O efeito Hawthorne leva esse nome em função da experiência homônima ocorrida em 1927 na fábrica *West Electric Company* (Chicago, EUA), onde se verificou que os indivíduos mudam seu comportamento de acordo com alterações percebidas no ambiente de trabalho. O efeito John Henry é assim chamado pois foi descoberto na *John Henry Company*, onde se introduziu, ainda em fase de testes, uma nova ferramenta que visava aumentar a produtividade. Na ocasião, os trabalhadores que foram mantidos usando a ferramenta antiga se sentiram desafiados a trabalhar mais, de forma a se mostrarem tão bons quanto os demais.

FICHA PADRONIZADA

FICHA PADRONIZADA PARA AVALIAÇÃO E CONTRATAÇÃO DE IMPACTO

POPULAÇÃO-ALVO	
NOME DO PROJETO	
LOCALIZAÇÃO GEOGRÁFICA	
BREVE DESCRIÇÃO DO PROJETO COM SUA TEORIA DA MUDANÇA (descrição dos insumos, atividades produtos e resultados)	
OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL RELACIONADOS (https://nacoesunidas.org/pos2015/agenda2030/)	
PERGUNTAS QUE DESEJA RESPONDER COM A AVALIAÇÃO	
MÉTRICA(S) ESCOLHIDA(S)	
NÍVEL DE MEDIÇÃO UTILIZADO	<input type="checkbox"/> Básico <input type="checkbox"/> Nível 1 <input type="checkbox"/> Nível 2 <input type="checkbox"/> Nível 3
JUSTIFICATIVA DA ESCOLHA DO NÍVEL DE MEDIÇÃO E POSSÍVEIS LIMITAÇÕES	

PERÍODO DA MEDIÇÃO INICIAL		
PERÍODO DA MEDIÇÃO FINAL		
DESCRIÇÃO PLANO AMOSTRAL (tamanho amostral e indicação de quem são os grupos de controle e tratado – caso sua amostra seja dividida em sub-grupos, por favor descreva quantos grupos e quantos indivíduos dentro de cada grupo estão presentes. Exemplo: 30 salas com 30 alunos cada)	GRUPO TRATADO	
	GRUPO DE CONTROLE (e placebo, se aplicável)	
EMPRESA EXECUTORA DO PROJETO		
INVESTIDOR DE IMPACTO		
ENDEREÇO DA EMPRESA EXECUTORA DO PROJETO		
RESPONSÁVEL PELO PROJETO		
TELEFONE		
E-MAIL		
ÁREA ABAIXO PARA PREENCHIMENTO DO AVALIADOR INDEPENDENTE DE IMPACTO		
RESPONSÁVEL PELA ANÁLISE	DATA	LOGO OU SELO DO AVALIADOR

Insp_er METRICiS