

Documento de Trabalho nº 1/13

Avaliação do Impacto: Breve Introdução



FICHA TÉCNICA

Pedro Amaral

Camões, Instituto da Cooperação e da Língua, I.P.

Gabinete de Avaliação e Auditoria

Fevereiro 2013

Os Documentos de Trabalho do Camões, IP, são produzidos por funcionários e colaboradores do Camões, IP.

Os Documentos de Trabalho são publicados sob responsabilidade única dos seus autores e não refletem necessariamente a opinião e posição do Camões, IP.

Índice

1. INTRODUÇÃO	4
2. HISTÓRIA DA AVALIAÇÃO DO IMPACTO	4
3. DEFINIÇÃO DE IMPACTO	5
4. PORQUÊ AVALIAR OS IMPACTOS?.....	7
5. RELAÇÕES DE CAUSALIDADE	9
6. ESTADO DA ARTE – ABORDAGENS E MÉTODOS.....	9
7. COMO AVALIAR OS IMPACTOS?.....	10
7.1.1. Abordagem experimental: Testes controlados aleatoriamente	13
7.1.2. Método de Comparação de resultados tendenciais	15
7.1.3. Diferença na diferença ou Diferença Dupla	15
7.1.4. O método das Variáveis Instrumentais ou Estimador das Variáveis Instrumentais	16
7.1.5. Séries Temporais Interrompidas	16
7.1.6. Identificação de processos	17
7.1.7. Modelo de Eliminação Geral	17
7.1.8. Análise da Contribuição.....	19
7.1.9. Técnica Participativa de Reporte da História do Desempenho.....	19
8. CONCLUSÃO.....	20
BIBLIOGRAFIA.....	21

1. INTRODUÇÃO

A avaliação do impacto é um dos aspetos mais importantes da avaliação da ajuda ao desenvolvimento e a definição das suas características, abordagens e metodologias assim como do seu âmbito e limitações têm sido um dos temas de avaliação mais discutidos em termos internacionais.

O presente documento tem como objetivo descrever resumidamente as características principais da avaliação do impacto assim como as principais abordagens e metodologias que permitem aferir o impacto de uma intervenção e as suas limitações. Procurou-se também contextualizar esta temática e esclarecer algumas questões relacionadas com ela, como as relações de causalidade. Os métodos apresentados não são os únicos métodos que podem ser aplicados numa avaliação do impacto mas apenas aqueles que mais têm sido discutidos na comunidade internacional.

2. HISTÓRIA DA AVALIAÇÃO DO IMPACTO

Depois de décadas de ajuda ao desenvolvimento, os principais problemas que afetam os países beneficiários dessa ajuda persistem: níveis elevados de pobreza, bem-estar reduzido, condições precárias de saúde e de educação, insegurança alimentar. Perante estes grandes problemas, as intervenções implementadas pelos doadores para lhes dar resposta tornaram-se cada vez mais abrangentes e complexas.

A constatação de que, perante as evidências disponíveis, era difícil determinar até que ponto as intervenções de desenvolvimento estavam a contribuir para alterar as condições de vida nos países em desenvolvimento fez crescer o ceticismo em relação à verdadeira eficácia da ajuda ao desenvolvimento. Tornou-se central responder à questão relacionada com saber o que funciona nessas intervenções e porque funciona.

A par desta questão, a preocupação com as grandes questões da gestão contemporânea: a responsabilização, a tendência para a gestão centrada nos resultados e o movimento para a tomada de decisão baseada em evidências, alertou igualmente para as lacunas existentes ao nível da informação necessária para responder a essa questão. A título de exemplo, o documento publicado em 2006, *“When Will We Ever Learn”*, pelo *Center for Global Development*, identificou uma lacuna de avaliação ao nível da ajuda ao desenvolvimento e apelou por mais e melhores avaliações do impacto rigorosas.

Tudo isto contribuiu para o aumento da necessidade de identificação dos resultados das intervenções e do seu contributo para o desenvolvimento dos países beneficiários da ajuda. É por este motivo que se realizaram uma série de eventos que colocaram a eficácia da ajuda no centro das preocupações da comunidade internacional, como, por exemplo, a aprovação dos Objetivos de Desenvolvimento do Milénio, em 2001, a Conferência de Monterrey sobre Financiamento do Desenvolvimento, em 2002, a Declaração de Paris sobre Eficácia da Ajuda de 2005, a Agenda para a Ação de Acra, em 2008 e a Declaração de Busan, em 2011.

Os países doadores deixaram de ficar satisfeitos com a constatação de que as suas intervenções estavam a produzir os outputs previstos para passar a ficar preocupados com a identificação e verificação do impacto que a ajuda estaria a ter na vida das populações-alvo.

Daí que a avaliação do impacto passasse a ser reconhecida como uma das principais áreas da avaliação da ajuda ao desenvolvimento.

Esta preocupação com o impacto da ajuda no desenvolvimento provocou a criação de diversas iniciativas que procuram desenvolver as abordagens e metodologias relacionadas com a avaliação do impacto através da elaboração de guias e documentos de trabalho e incentivar a sua aplicação prática. Entre as principais iniciativas relacionadas com esta temática podemos incluir a NONIE1 (*Networks of Networks on Impact Evaluation*), criada em 2007, e a *International Initiative for Impact Evaluation*² (3ie). Além destas iniciativas, algumas organizações internacionais como o Banco Mundial, a ONU e a Comissão Europeia revelaram uma atenção redobrada acerca da avaliação do impacto.

3. DEFINIÇÃO DE IMPACTO

O impacto é considerado como o ultimo elo na chamada cadeia de resultados, que relaciona os *inputs* de uma intervenção de desenvolvimento com os seus resultados de médio e longo prazo (Fig 1).

¹ Constituída pela Rede de Avaliação do CAD/OCDE, o Grupo de Avaliação das Nações Unidas (UNEG), o Evaluation Cooperation Group (ECG) e pela International Organization for Cooperation in Evaluation (IOCE)

² Constituída por diversas agências doadoras bilaterais, instituições financeiras internacionais e instituições dos países parceiros.

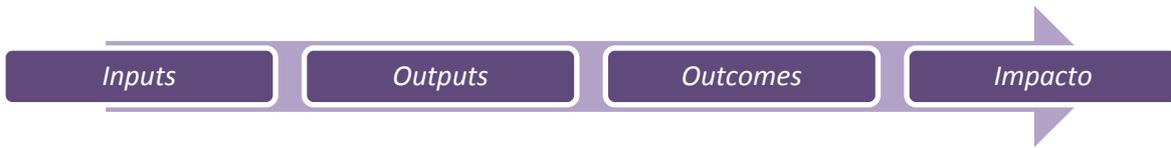


Figura 1 – Cadeia de Resultados

O impacto de uma intervenção de desenvolvimento consiste nos efeitos resultantes da implementação dessa intervenção num determinado local, ao nível dos indicadores sociais, económicos, ambientais, entre outros e das alterações comportamentais nos beneficiários finais.

Assumido como um dos critérios de avaliação da ajuda ao desenvolvimento pelo CAD/OCDE, o impacto é definido, de forma abrangente, no glossário desta organização como sendo os *“efeitos de longo prazo, tanto positivos como negativos, primários e secundários, produzidos por uma intervenção de desenvolvimento, previstos ou não”*.

Entendido desta forma, como afirma Van der Berg, o impacto *“não é definido como uma relação mas como um tipo de estado final ou uma fotografia dos efeitos passado bastante tempo depois do fim da intervenção”* (Berg, 2011:11). Ora, impacto tem vindo a ser interpretado cada vez mais como uma relação, a relação causal entre as ações de uma determinada intervenção e as alterações provocadas por estas nos indicadores e nos comportamentos dos beneficiários finais. São estas relações causais que estão no centro da avaliação do impacto.

Em termos gráficos, o impacto pode ser representado da seguinte forma:

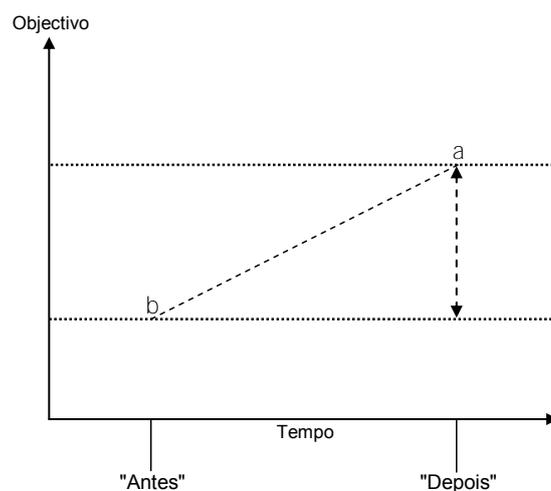


Figura 2- Representação gráfica do impacto de uma intervenção

Assim, podemos dizer, simplisticamente, que antes de a intervenção ter começado a ser implementada a situação era **b** (que funcionará como *baseline*) e que após esta ter sido implementada, a situação passou a ser **a**, pelo que, o impacto da intervenção é a diferença entre **b** e **a** (seta a tracejado).

4. PORQUÊ AVALIAR OS IMPACTOS?

Avaliar o impacto de uma determinada intervenção de desenvolvimento não se limita a verificar o que aconteceu mas também verificar porque é que isso aconteceu. Ao determinar os efeitos de uma intervenção, a avaliação do impacto está a contribuir diretamente para a responsabilização e prestação de contas, tanto na dimensão ascendente como na sua dimensão descendente, ou seja, tanto em relação aos financiadores, doadores e cidadãos, assegurando que as suas contribuições são aplicadas em intervenções eficazes, como em relação aos beneficiários e comunidades locais, ao informá-los sobre se a intervenção está a beneficiá-los ou não.

A avaliação do impacto também contribui diretamente para a tomada de decisão acerca do futuro da intervenção e da sua possível aplicação/transposição para outro contexto ou aumento da sua dimensão (*scale up*), na medida em que verifica aquilo que funciona e aquilo que não funciona e se as características que funcionam podem ser transpostas ou redimensionadas.

Ao identificar os processos de mudança que levaram aos impactos observados, a avaliação do impacto contribui também para a aprendizagem e criação de conhecimento sobre a cooperação internacional e a avaliação. Nesta linha, potencia igualmente a aprendizagem organizacional, já que permite que a organização tenha consciência daquilo que funciona, ou não, e porquê.

Resumindo, a avaliação do impacto preocupa-se essencialmente com as seguintes questões:

- Determinar o alcance e a amplitude dos *outcomes* de uma intervenção;
- Determinar se uma intervenção está a ser implementada como planeado e como é que a intervenção influenciou os seus *outcomes*;
- Dar evidências aos financiadores, gestores e políticos sobre em que medida os recursos alocados foram gastos de forma eficiente;
- Informar a tomada de decisão acerca da possibilidade da sua transposição para outros contextos ou do alargamento da sua dimensão.

A questão da atribuição é um dos aspetos da avaliação do impacto que tem recebido muita atenção nos últimos anos. Saber se a intervenção causou de facto os impactos observados,

isolando e estimando o contributo particular da intervenção, é uma das questões que pode ser respondida por uma avaliação do impacto mas não deve sobrepor-se a outras, igualmente importantes. Não chega saber se a intervenção produziu ou não os impactos, é igualmente importante, por exemplo, compreender de que forma a intervenção os produziu e quais os fatores que lhe permitiram ter sucesso, ou não.

De forma geral, as principais dimensões que envolvem a avaliação do impacto podem ser resumidas como no quadro 1.

Quadro 1 – Avaliação do impacto

Dimensão	Propriedades
Orientação	Estabelecimento do valor do programa Justificação das decisões para aumentar o programa Prestação de contas perante os financiadores e outros detentores de interesse.
Principais Questões	O programa foi implementado como planeado? Os objectivos definidos do programa foram alcançados? As necessidades dos que beneficiaram do programa foram satisfeitas? Quais são os <i>outcomes</i> não pretendidos? A estratégia de implementação leva a <i>outcomes</i> não pretendidos? Como é que as diferenças na implementação afectam os <i>outcomes</i> do projecto? Dados os custos do programa quais são os seus benefícios?
Estado do programa	Implementado
Principal Enfoque	Centra-se na implementação e/ou <i>outcomes</i> . Os estudos mais abrangentes combinam tanto a implementação como os <i>outcomes</i> , conhecidos como estudos processo- <i>outcomes</i> .
Timing (relativo à implementação do programa)	Nominalmente “após” o programa ter completado pelo menos um ciclo com os beneficiários do programa. Na prática, os estudos de impacto devem ser realizados a qualquer altura depois do programa estar “implementado”.
Principais abordagens	Baseada nos objectivos Baseada nas necessidades Sem objectivos Estudos de processo- <i>outcome</i> Avaliação realista Auditoria do desempenho Avaliação baseada na teoria
Recolha de evidências	Utilização de testes e dados quantitativos, sempre que exequível a utilização de grupos de tratamento e de controlo. Os estudos de implementação exigem geralmente dados de observação. Determinar todos os <i>outcomes</i> requer a utilização de métodos mais exploratórios e a utilização de evidências qualitativas.

Adaptado de Owen, 1999

5. RELAÇÕES DE CAUSALIDADE

Como referido, a principal característica da avaliação do impacto é o facto desta, mais do que tentar identificar e documentar os potenciais impactos de uma intervenção, procurar compreender como é que esses impactos foram produzidos, identificando qual a relação causal entre estes e a intervenção, ou seja, inferindo a causalidade.

Inferir a causalidade significa ligar uma causa a um efeito, estabelecer uma relação causal entre dois fenómenos, neste caso a intervenção e o impacto. Nesta análise é preciso ter em atenção que existem diversos tipos de relações causais e diferentes abordagens à inferência da causalidade (ver bibliografia). Além disso, importa referir que, para se considerar um fenómeno como a causa de outro, este deve ser necessário e/ou suficiente para o produzir. Neste processo de inferência causal, devem ser consideradas e eliminadas todas as explicações alternativas ou rivais.

Determinar a causalidade é uma questão fundamental (não só) da avaliação do impacto, já que permite distinguir entre os *outcomes* ou impactos de uma intervenção, que são causados por esta, daquilo que não passa de uma coincidência, porque não foi provocado pela intervenção.

Tendo em conta que as intervenções de desenvolvimento nem sempre são intervenções simples, em que apenas existe uma única cadeia causal, mas antes intervenções em que se verificam diversas cadeias causais simultâneas que produzem os impactos e em que existem diversos mecanismos causais em funcionamento, a avaliação do impacto torna-se mais complexa quando comparada com outros tipos de avaliação, ao mesmo tempo que se depara com este desafio da complexidade.

6. ESTADO DA ARTE – ABORDAGENS E MÉTODOS

A causalidade, e por consequência, o impacto, pode ser determinada através de várias abordagens e metodologias. Não existe uma metodologia específica para se avaliar o impacto. As metodologias que são utilizadas para aferir o impacto são comuns a outras áreas e temáticas. Exemplo disso é a análise custo-benefício ou a análise custo-eficácia, as quais são utilizadas tradicionalmente para avaliar a eficiência de uma intervenção mas que são também ferramentas excelentes para avaliar o seu impacto.

Apesar da diversidade metodológica existente para avaliar o impacto, a questão de definir o que é que constitui uma evidência credível e sólida da causalidade entre intervenção e impacto e de

como se obtém essa evidência provocou um aceso debate na comunidade internacional do desenvolvimento. Resumidamente, a questão está em saber se há uma abordagem *gold standard* para a avaliação do impacto, ou seja, se existe apenas uma única abordagem rigorosa para aferir os impactos. Foi a intenção de impor a abordagem experimental ou testes controlados aleatoriamente (descritos mais adiante) como *gold standard* deste tipo de avaliação que suscitou este debate. Não está em causa a validade e o valor destes testes mas sim o facto de existirem outras abordagens que permitem igualmente inferir a causalidade e, conseqüentemente, determinar o impacto. Como reconhece Michael Scriven, um dos principais críticos desta abordagem, estes testes, “(...) nos últimos 30 anos, não têm sido utilizados tão frequentemente como deveriam ter sido” (Scriven, 2006: 8).

Se é verdade que o facto de este debate revelar uma grande preocupação com a procura de um conhecimento mais profundo sobre o que funciona e não funciona no desenvolvimento é um fator positivo, essa procura não se deve limitar apenas a um dos vários métodos disponíveis, como afirma Martin Ravallion (2009):

“A importância que os investigadores estão atualmente a dar à obtenção de um melhor conhecimento acerca da eficácia do desenvolvimento é bem-vinda. A seleção aleatória é uma das ferramentas que pode ajudar. Contudo, a tarefa importante na investigação sobre aquilo que funciona e aquilo que não funciona na luta contra a pobreza não pode ser monopolizada por um método”.

O que é importante na escolha das metodologias é ter em atenção que, como refere Howard White (White, 2009: 13), não existe uma hierarquia de métodos. A solução poderá passar pela utilização destes métodos em conjunto, os chamados métodos mistos (*mixed methods*), tal como proposto por Jennifer Green, já em 1989 (Green, 1989) de forma a potenciar os resultados de uma avaliação do impacto.

7. COMO AVALIAR OS IMPACTOS?

O processo de seleção da abordagem e dos métodos a utilizar deve ponderar cuidadosamente um conjunto de aspetos relacionados com a avaliação do impacto e que se relacionam com a qualidade e credibilidade do seu produto final.

A avaliação do impacto não é exclusiva de nenhum método específico. Pelo contrário, são vários os métodos que permitem inferir a causalidade e determinar o impacto de uma intervenção. O

essencial é que, independentemente da abordagem adotada para se proceder à avaliação, no momento da seleção dos métodos a utilizar, se tenha em atenção se estes são **úteis, credíveis, exequíveis e éticos**, ou seja, devem ser adequados à situação, já que situações diferentes exigem tipos diferentes de avaliação do impacto. Além disso, as expectativas dos detentores de interesse e dos potenciais utilizadores da avaliação devem ser, igualmente, tidas em conta. Neste sentido, essa seleção deve ser determinada por:

- **Finalidade da avaliação:** por exemplo, se aquilo que se pretende for a responsabilização pelos resultados verificados, será suficiente utilizar uma abordagem que infira a relação causal. Por sua vez, se o propósito for promover a aprendizagem e o conhecimento, então será necessário utilizar uma abordagem que explique como e porque é que se alcançaram os impactos verificados.
- **Perguntas da avaliação:** é fundamental escolher os métodos que melhor respondam às perguntas da avaliação, podendo aplicar-se um método específico a cada uma das perguntas. Devem ser as perguntas da avaliação a determinar os métodos a utilizar e não estes a determinar as perguntas.
- **Recursos disponíveis:** os recursos financeiros, humanos, as questões logísticas e os prazos estipulados devem influenciar a escolha da abordagem, uma vez que algumas abordagens exigem a disponibilidade de recursos consideráveis, como é o caso da abordagem experimental.
- **Natureza da intervenção:** se a natureza da intervenção for padronizada ou adaptativa, isso tem implicações no método a escolher, visto que nem todos os métodos são adequados para avaliar uma intervenção adaptativa.

É necessário referir que a combinação de abordagens e métodos diferentes poderá potenciar a sua eficácia, na medida em que aproveita as potencialidades e sinergias de cada um deles e contribui para ultrapassar algumas das suas limitações. Esta utilização combinada foi bastante facilitada pelo novo paradigma desenvolvido nos últimos anos na área das ciências sociais, como afirmam os autores citados em Stern (2012), o qual rejeita a divisão simples entre abordagens quantitativas e qualitativas, promovendo assim o esbatimento das fronteiras de separação entre essas duas abordagens e a sua utilização conjunta.

Quadro 2 – Abordagens, Variantes Específicas e Bases para Inferência Causal

Abordagens	Variantes específicas	Bases para inferência causal
Experimental	Testes Controlados Aleatoriamente; Quase Experimental; Experimentações naturais.	Contrafactual; a presença em simultâneo da causa e efeito.
Estatística	Modelação Estatística; Estudos longitudinais; Econometria.	Correlação entre causa e efeito ou entre variáveis, influência de causas múltiplas (normalmente) isoláveis num único efeito. Controlo das "variáveis de confusão".
Baseada na teoria	<i>Abordagens do processo causal</i> : Teoria da Mudança, identificação de processos, Análise da Contribuição, Percurso do impacto; <i>Abordagens do mecanismo causal</i> : Avaliação Realista, Análise de Congruência.	Identificação/confirmação de processos ou "cadeias" causais; Factores de suporte e mecanismos intervenientes no contexto.
Baseada no "caso"	<i>Interpretativa</i> : Naturalista, Grounded Theory (Teoria fundamentada nos dados), etnografia; <i>Estruturada</i> : Configurações, Análise Qualitativa Comparativa, Análise individual de cada caso, Simulações e análise de redes.	Comparação cruzada e individual dos casos de combinações de factores causais; Generalização analítica baseada na teoria.
Participativa	<i>Abordagens normativas</i> : Avaliação participativa ou democrática, avaliação de <i>empowerment</i> <i>Abordagens de agência</i> : Aprender fazendo, Diálogo político, Pesquisa-Ação Colaborativa.	Validação pelos participantes de que as suas acções e os efeitos experimentados são "causados" pelo programa; Adopção, personalização e compromisso com uma meta.

Fonte: (Stern, 2012)

No quadro 2 foram apresentadas, a título de exemplo, as principais abordagens à avaliação do impacto. Descreve também as principais variantes de cada uma dessas abordagens e os métodos que podem ser utilizados. Finalmente, apresenta as bases a partir das quais cada uma dessas abordagens infere a causalidade.

São apresentados e descritos de forma sucinta, nas próximas páginas, alguns dos métodos referidos no quadro anterior e que permitem aferir o impacto de uma intervenção. Não se faz uma descrição exaustiva de todos os métodos disponíveis que permitem avaliar o impacto mas procura-se que os métodos apresentados e descritos abranjam o máximo das abordagens existentes para avaliar o impacto. Procura-se, igualmente, apresentar métodos mais recentes, como a técnica participativa de reporte da história do desempenho, de forma a dar uma ideia das potencialidades que ainda existem para o desenvolvimento da avaliação do impacto. A consulta da bibliografia pode ajudar a encontrar informação adicional sobre estes e outros métodos existentes.

7.1.1. Abordagem experimental: Testes controlados aleatoriamente (Randomized Controlled Trials)

Os métodos experimentais, ou testes controlados aleatoriamente, pretendem apreciar os impactos de uma intervenção utilizando um modelo metodológico importado da ciência agrária e da medicina, daí serem também chamados de estudos laboratoriais (Owen, 1999: 267), visto que pretendem alcançar um nível de rigor científico nas suas conclusões. Um dos conceitos fundamentais desta metodologia é o conceito de **contrafactual**, que é a estimativa de qual seria a situação dos participantes, em termos de indicadores de *outcomes*, caso a intervenção não tivesse sido implementada. A análise do contrafactual baseia-se na comparação entre esta situação e a situação factual, ou seja, a situação real, que beneficiou da intervenção. É a comparação entre estas duas situações que torna possível verificar o impacto dessa intervenção.

A identificação da situação contrafactual é feita através da criação de dois grupos pelos quais se distribuem os sujeitos ou unidades de análise (que podem ser indivíduos, grupos de indivíduos ou organizações como escolas, hospitais, etc.): o **grupo experimental** (de estudo ou de tratamento) e o **grupo de controlo** (ou de comparação). Os sujeitos que são incluídos no grupo experimental irão beneficiar da intervenção enquanto os sujeitos incluídos no grupo de controlo não beneficiarão dela. Este grupo corresponde, desta forma, à situação contrafactual. A diferença entre os *outcomes* médios de cada um dos grupos, obtida através da sua respetiva comparação, será o impacto da intervenção.

A questão fundamental é garantir que o impacto verificado se deve à intervenção e não a outros fatores. Para isso é essencial garantir que os sujeitos que irão fazer parte dos grupos possuam todos as mesmas hipóteses, *ex-ante*, de participar na intervenção. Devem ser, também, estatisticamente equivalentes, sendo os seus atributos observáveis e não observáveis distribuídos de forma igual pelos dois grupos. A esta seleção de sujeitos, chama-se **seleção aleatória** (*randomization*), a qual, ao distribuir aleatoriamente sujeitos semelhantes pelos dois grupos, garante que a única diferença entre estes seja o facto de uns terem recebido a intervenção e os outros não. Se a única diferença entre esses grupos é a participação ou não na intervenção, então a diferença que se verificar nos indicadores de *outcome*, após a intervenção, ou seja, o impacto, deve-se unicamente à intervenção. A seleção aleatória dos sujeitos, ao eliminar o enviesamento de seleção (*selection bias*), contribui para que seja possível atribuir o impacto a uma dada intervenção, contribuindo para a sua elevada validade interna.

Apesar da sua elevada validade interna, esta abordagem apresenta algumas ameaças que podem por em causa os resultados das suas experimentações, como se pode verificar nos exemplos descritos na próxima caixa.

Ameaças à validade interna da análise de contrafactual/testes controlados aleatoriamente

Sobredeterminação: surge em situações em que para um mesmo efeito existem duas ou mais causas (minimamente) suficientes e distintas, o que coloca em causa a análise do contrafactual, que se baseia numa causalidade linear, uma causa, um efeito.

Efeito Hawthorne: este efeito consiste na tendência existente de as pessoas incluídas numa investigação reagirem, positiva ou negativamente, ao próprio facto de estarem a ser investigadas, alterando por isso o seu comportamento. Isto dificulta a identificação e atribuição dos impactos de uma intervenção.

Relacionado com este efeito, se bem que diferente, existe também o **efeito placebo**.

Contaminação: situação em que algo que não a intervenção tem um efeito não detetado nos efeitos dessa intervenção. A contaminação pode ocorrer tanto a nível interno, isto é, entre os dois grupos, em que o comportamento de um grupo influencia o outro, e a nível externo, visto ser difícil isolar estas unidades de análise de outras intervenções que podem produzir alterações nos indicadores de impacto.

Desgaste diferencial (*differential attrition*): relaciona-se com a possibilidade de os sujeitos abandonarem qualquer um dos grupos durante o período de análise, o que prejudica a comparabilidade e equivalência dos grupos. Este desgaste torna-se diferencial porque não é possível assumir que os sujeitos que faltam num grupo possuam as mesmas características relevantes para os *outcomes* que os sujeitos que faltam no outro.

Efeito caixa negra (*black box*): apesar destes testes poderem conseguir identificar a relação causal entre intervenção e efeito, não conseguem explicar porque e como é que isso aconteceu, o que acaba por limitar o alcance da análise e pode prejudicar a validade externa desta abordagem.

Experimentação Prematura: esta abordagem aplica uma metodologia utilizada no domínio da investigação clínica de forma a alcançar o mesmo rigor, só que nesse domínio, os testes controlados aleatoriamente são apenas uma das quatro fases da experimentação. Aplicar apenas uma dessas fases não é suficiente para se efetuar uma apreciação científica, só com a realização das quatro fases isso é possível.

Validade externa: apesar de ambicionarem ser fortes em termos de validade interna, estes testes possuem uma validade externa reduzida, dificultando bastante a sua generalização, provavelmente o que mais interessa aos decisores políticos.

Resta salientar as questões éticas que a aplicação da abordagem experimental suscita, uma vez que impedir deliberadamente um conjunto alargado de pessoas de beneficiar de uma intervenção que poderia melhorar a sua qualidade de vida, satisfazendo as suas necessidades mais básicas é

uma questão complexa e de difícil justificação. Como afirma Martin Ravallion, “Pensar o desenvolvimento como um “tratamento” é bastante infeliz (...)” (Ravallion, 2009).

7.1.2. Método de Comparação de resultados tendenciais (Propensity Score Matching)

A **análise de correspondência** (*matching*) tem como base a utilização de técnicas estatísticas que lhe permitem construir um grupo de controlo artificial através da identificação dos sujeitos que, dentro do grupo de sujeitos que não participaram na intervenção, são o mais semelhantes possível com os que participaram e que são o grupo experimental. Essa semelhança é aferida através da identificação das características individuais observáveis nesse grupo de sujeitos. A questão está em que se forem poucas características, fazer essa correspondência é fácil, o que raramente é o caso.

A solução para esta situação é o chamado método de **comparação de resultados tendenciais** (*propensity score matching*). Este método não necessita de fazer corresponder todos os sujeitos que participam na intervenção com todos os sujeitos que não participam. Em vez disso, a comparação das variáveis observáveis é feita com base na semelhança da sua probabilidade de virem a participar na intervenção. Esta probabilidade é aquilo a que se chama de **resultado tendencial** (*propensity score*). O grupo de comparação é constituído pelo grupo de sujeitos não participantes no programa que possuem o resultado tendencial mais próximo dos que fazem parte do grupo experimental, constituindo-se, desta forma, como o contrafactual. O impacto é estimado através da comparação da diferença entre os *outcomes* médios finais do grupo experimental e os *outcomes* médios finais do grupo de comparação cuja correspondência com o grupo experimental foi feita estatisticamente.

A diferença entre este método e o método experimental é o facto de este se basear tanto nas características observáveis como não observáveis dos sujeitos, enquanto o método de comparação de resultados tendenciais apenas se baseia nas características observáveis. Por este motivo, a sua realização exige a disponibilidade de informação de partida fidedigna.

7.1.3. Diferença na diferença (Difference-in-difference) ou Diferença Dupla (Double Difference)

Este método procura construir a situação contrafactual, comparando a mudança, ao longo do tempo, nos *outcomes* observados do grupo experimental e do grupo de comparação, antes e depois de a intervenção ter sido implementada. Antes de a intervenção ser implementada, é

necessário existir informação de partida relativamente aos *outcomes* de cada um desses grupos. Depois da intervenção, reúne-se novamente essa informação e é feita a comparação da diferença verificada nos *outcomes* dos dois grupos. Mais do que fazer uma simples comparação antes e depois, este método compara duas diferenças (daí a sua designação). A primeira diferença é a diferença nos *outcomes* antes e depois da intervenção verificada no grupo experimental enquanto a segunda diferença é a mudança verificada, antes e depois, nos *outcomes* do grupo de comparação, que não recebeu a intervenção. Desta forma, este método identifica as diferenças, constantes no tempo, entre estes grupos, sendo o impacto da intervenção medido através do cálculo da diferença entre as duas diferenças referidas.

7.1.4. O método das Variáveis Instrumentais (*Instrumental Variables Method*) ou Estimador das Variáveis Instrumentais (*Instrumental Variables Estimator*)

Este método é uma variável do **método de regressão**. Considera que o processo de colocação na intervenção é endógeno a essa intervenção, relacionado com os *outcomes* da intervenção. Nestas condições, a fim de evitar o enviesamento de análise se se aplicasse o método normal de regressão, este método procede à substituição da variável que determina a participação ou não no programa por uma **variável instrumental**. Esta variável deve imitar a variável que substitui, estando a diferença entre estas no facto de que esta variável independente não estar diretamente relacionada com o *outcome* pretendido da intervenção. A título de exemplo, numa intervenção de microcrédito que distribua folhetos informativos de forma aleatória em algumas localidades, como a leitura do folheto pode levar à participação na intervenção mas não afecta diretamente o seu *outcome*, a variável instrumental dessa participação pode ser o facto de esses folhetos terem sido ou não distribuídos numa dada localidade, o que permitirá comparar a diferença de *outcomes* entre localidades que receberam e não receberam os folhetos e aferir o impacto da intervenção.

7.1.5. Séries Temporais Interrompidas (*Interrupted Time Series*)

A série temporal é criada através da realização de um número significativo de observações no tempo, antes e depois da intervenção. Desta forma, é possível identificar a evolução nos indicadores de *outcomes* da intervenção, que são a sua variável dependente. Se o gráfico desses indicadores (variável) revelar uma mudança abrupta de nível ou de direção precisamente no ponto em que a intervenção foi introduzida, então é possível afirmar que a intervenção é a causa

desse efeito nos indicadores de *outcomes*. É desta forma que este método permite identificar o impacto da intervenção.

A principal preocupação desta metodologia é garantir que os efeitos observados se devem diretamente à intervenção. Ela consegue separar os efeitos reais da intervenção de tendências de longo prazo na série temporal através da realização de bastantes observações antes e depois da intervenção.

7.1.6. Identificação de processos (Process Tracing)

O método de identificação de processos permite analisar relações causais complexas entre uma causa potencial e um efeito observado. O seu objetivo é identificar os mecanismos ou processos causais que relacionam essa causa com esse objetivo. A análise dos *outcomes* é **retrospectiva**, já que recua no tempo, através dos principais acontecimentos, processos e decisões que caracterizam a cadeia causal, até chegar à causa do impacto. Com base no impacto observado, o investigador determina um conjunto de **hipóteses**, preferencialmente concorrentes entre si, as quais serão depois testadas através da construção e análise de uma **cronologia narrativa** que descreva a sequência de acontecimentos, construindo assim a cadeia causal. Uma vez construída a cadeia causal, testa-se a **congruência** ou incongruência entre as hipóteses que foram formuladas e aquilo que foi observado e analisado. Desta forma, é possível aferir o impacto de uma intervenção, identificando a intervenção como a sua causa efetiva.

Esta metodologia adapta-se especialmente a situações complexas, que envolvam dinâmicas de mudança, já que consegue captar a forma como a situação evoluiu. Além disso, consegue também descrever a forma como os atores sociais reagem no seu contexto, aos diversos estímulos das condições internas e externas da intervenção. Este método necessita de grandes quantidades de informação, proveniente de diversas fontes, de forma a evitar que a sua análise seja enviesada por falta de informação. Por este motivo, esta análise é complexa e exige tempo para analisar a toda a informação disponível de forma rigorosa.

7.1.7. Modelo de Eliminação Geral (General Elimination Model)

Esta abordagem metodológica (relacionada em parte com a metodologia descrita no ponto anterior) foi desenvolvida por Michael Scriven para a área da avaliação, incluindo avaliação do impacto. Constitui o fundamento da lógica básica subjacente a todas as metodologias que procuram fazer afirmações causais, como os testes controlados aleatoriamente. A sua premissa

geral é o princípio determinístico, que estipula que todos os acontecimentos macro possuem uma causa própria.

Partindo do acontecimento tipo que estamos interessados em analisar, como por exemplo, o declínio ou aumento significativo na taxa de escolarização num período x, constrói-se uma **lista de causas possíveis**. Esta lista de causas baseia-se no conhecimento adquirido pela prática de outros peritos/especialistas que investigaram aprofundadamente este tipo de acontecimento/efeito, o que permite identificar as suas possíveis causas. Sempre que se esteja perante novos efeitos, essa lista poderá ter que ser adaptada e alargada. Scriven dá o exemplo das listas de motivos e suspeitos, utilizadas no âmbito da investigação criminal.

Uma vez construída a lista de causas possíveis, é necessário analisar o seu **modus operandi**, ou seja, verificar se, nessa causa, está presentes a sequência ou cadeia de acontecimentos intermédios e concorrentes que teria que estar presente se essa causa fosse a causa real do efeito observado. No exemplo do detetive, seria a lista dos meios e oportunidades que cruzaria com a lista dos suspeitos.

Finalmente, com base na verificação da existência ou não de fatores que foram listados com a análise *modus operandi* de cada uma das causas possíveis, é altura de analisar os **factos do caso** e eliminar as causas cujos *modus operandi* não estejam completamente presentes. Eventualmente, as causas que resistam a esta análise serão as causas reais do efeito observado.

“Para dar um exemplo de um trabalho em que tenho estado envolvido, ao olhar para o efeito da ajuda prestada por Heifer ou Gates a agricultores extremamente pobres na África Oriental, depois de determinar que a seguir ao início da ajuda se seguiu uma melhoria substancial no bem-estar, a qual foi sustentável durante alguns anos, nós procurámos a presença de mais de uma dúzia de outras causas possíveis para esse aumento subsequente observado no bem-estar, incluindo: os esforços do governo local que desceu até ao nível da aldeia, os esforços análogos de outros filantropos, os ganhos de autoajuda resultantes de uma liderança inspirada nas comunidades locais, os rendimentos melhorados de membros das famílias que viajam para locais com emprego bem remunerado noutros lados e que remetem esse dinheiro para casa, o aumento dos preços de leite ou dos bezerros nos mercados locais, os resultados benéficos de alguns anos de bom tempo ou da melhoria do abastecimento de água ou das melhorias motivadas pela tecnologia na qualidade de alimentação comercial, medicamentos veterinários ou sementes melhoradas de erva para pastagens. Isto requer um esforço sistemático considerável (...)” (Scriven, 2008).

Como afirma Scriven, esta abordagem “(...) é essencialmente uma extensão do senso comum”. (Scriven, 2008). Desde que esta análise seja rigorosamente aplicada, as suas conclusões possuem o mesmo nível de confiança que, por exemplo, as conclusões formuladas por uma abordagem experimental. Esta abordagem é a mesma que é utilizada em áreas do conhecimento como a medicina legal, epidemiologia, mecânica, investigação criminal, história.

7.1.8. Análise da Contribuição

Esta ferramenta analítica, originária da área da gestão financeira e das análises na perspetiva do negócio, foi desenvolvida por John Mayne como forma de resolver a questão da atribuição na avaliação. Aplicando-se a intervenções que se estruturam numa **teoria da mudança**, esta abordagem pretende ser alternativa a outras abordagens, como a experimental, nas situações em que essas não sejam viáveis.

Nas palavras de Mayne, a análise da contribuição é uma “(...) análise específica realizada para proporcionar informação sobre a contribuição de um programa nos *outcomes* que este está a tentar influenciar.” (Mayne, 1999: 6).

Num contexto em que o estabelecimento da causalidade, como já referido, é complexo, esta metodologia considera que o desafio não é tanto demonstrar a atribuição, ou seja, provar que a intervenção foi a causa daquele impacto, mas demonstrar a contribuição, ou seja, estabelecer a relação entre a intervenção e impacto, através da recolha de evidências de que a intervenção contribuiu para o referido impacto. A causalidade é analisada através da verificação da teoria da mudança subjacente à intervenção e de outros fatores que a possam ter influenciado.

Apesar de esta metodologia ser bastante útil para apreciar os impactos de uma intervenção, ela depende da informação rigorosa proveniente de um sólido sistema de acompanhamento do desempenho.

7.1.9. Técnica Participativa de Reporte da História do Desempenho (Participatory Performance Story Reporting Technique)

Esta metodologia assenta principalmente num tipo específico de relatório de **história do desempenho**. Esta história consiste num sumário do desempenho de uma intervenção. Procura explicar não só aquilo que o programa alcançou de facto mas também apresentar as relações causais que demonstram a forma como os resultados foram alcançados. Para isto baseia-se numa cadeia de resultados ou na teoria do programa, que hierarquizam os *outcomes*.

Desenvolvida por Jessica Dart (a partir de elementos da análise da contribuição), esta técnica caracteriza-se por possuir um processo participativo de cinco etapas que gera a informação. O relatório, que apresenta essa informação, uma vez construído, apresenta uma estrutura, também constituída por cinco partes diferentes. Estes relatórios, ao analisarem a forma como a intervenção contribuiu para os *outcomes*, acabam por ser uma ferramenta útil para apreciar os seus impactos.

A informação que vai preencher este relatório é gerada através do referido **processo participativo**, que consiste numa oportunidade fundamental para a aprendizagem e reflexão de todos os detentores de interesse que venham a ser envolvidos no processo.

Esta metodologia deve ser aplicada em conjunto com outras metodologias, de forma a evitar alguns dos riscos que apresenta em termos de enviesamento da análise. O facto de se focar apenas numa aspeto da avaliação, a questão da contribuição, e não em outras questões, como a do custo-eficácia da intervenção, faz com que o seu âmbito possa ser algo limitado. Apesar disto, acaba por ser uma ferramenta sofisticada e fácil de utilizar para apreciar os impactos de uma intervenção.

8. CONCLUSÃO

Avaliar os efeitos de médio e longo prazo das intervenções nos beneficiários finais é uma das principais tendências atuais no domínio da avaliação da ajuda ao desenvolvimento. A procura de evidências credíveis acerca do que funciona, em que circunstâncias e para quem nas intervenções de desenvolvimento, tornou a avaliação do impacto num dos tipos de avaliação mais discutidos internacionalmente e um dos mais importantes para a prática dessa disciplina.

São várias as abordagens e métodos com potencialidades para estabelecer a referida relação de causalidade entre intervenção e impacto. Esta questão torna a avaliação do impacto uma questão complexa, em que à necessidade de rigor acresce a exigência dos seus resultados serem credíveis. A decisão de avaliar o impacto não é linear, implica ter em conta um conjunto alargado de questões relacionadas com o tipo de intervenção a avaliar, o contexto em que esta se desenvolveu e a própria finalidade da avaliação.

A avaliação do impacto não é ainda uma prática corrente no âmbito da cooperação portuguesa. Nesse sentido, a descrição dos métodos pretende contribuir para a sua divulgação, disponibilizando, a todos os interessados, informação básica sobre cada um deles. Pretende, ao

mesmo tempo, constituir um ponto de partida para uma análise mais aprofundada acerca das abordagens e métodos para avaliação dos impactos e, desta forma, aprofundar o seu conhecimento, potenciando, a prazo, a realização de avaliações do impacto na cooperação portuguesa.

BIBLIOGRAFIA

- ASIAN DEVELOPMENT BANK (2006), Impact Evaluation: Methodological and Operational Issues.
- BARAHONA, Carlos (2010), Randomised Control Trials for the Impact Evaluation of Development Initiatives: A Statistician's Point of View, *ILAC Working Paper*, nº 13, International Learning and Change (ILAC) Initiative.
- BENNET, Andrew e GEORGE, Alexander L. (1997), *Process Tracing in Case Study Research*, McArthur Foundation Workshop on Case Study Methods.
- BRASS, Clinton T., NUNEZ-NETO, Blas e WILLIAMS, Erin D. (2006), *Congress and Program Evaluation: An Overview of Randomized Controlled Trials (RCTs) and Related Issues*, CRS Report for Congress.
- CHAMBERS, Robert [et a.] (2009), Designing impact evaluations: different perspectives, International Initiative for Impact Evaluation, *Working Paper 4*.
- COOK, Thomas D. [et al.] (2010), Contemporary Thinking About Causation in Evaluation: A Dialogue With Tom Cook and Michael Scriven, *American Journal of Evaluation*, 31(1): 105-117.
- DART, Jessica (2008), *Report on outcomes and get everyone involved: The Participatory Performance Story Reporting Technique*, Apresentação feita na conferência da AES em Perth.
- DAVIDSON, E. Jane (2004), *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*, Sage Publications.
- DAVIDSON, E. Jane (2009), *Causal inference: Nuts and bolts*, Mini Workshop for the Anzea Wellington Branch.
- DONALDSON, Stewart, CHRISTIE, Christina A. (2004), The 2004 Claremont Debate: Lipsey vs. Scriven. Determining Causality in Program Evaluation & Applied Research: Should Experimental Evidence be the Gold Standard?, *Journal of Multidisciplinary Evaluation*, nº 3: 60-77.
- DONALDSON, Stewart, CHRISTIE, Christina A. e MELVIN, Mark M. (2008), *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Thousand Oaks, CA, Sage.
- DYBDAL, Line [et.al.] (2011), Contribution Analysis Applied: Reflections on Scope and Methodology, *The Canadian Journal of Program Evaluation*, Vol. 25, Nº 2: 29-57.
- FALLETI, Tulia G. (2006) Theory Guided Process-Tracing in Comparative Politics: Something Old and Something New, *Newsletter of the Organized Section in Comparative Politics of the APSA*, APSA-CP, nº 1.
- FORSS, Kim, MARRA, Mita, SCHWARTZ, Robert (2011), *Evaluating the Complex: Attribution, Contribution, and Beyond*, Transaction Publishers.
- FUNNEL, Sue C., ROGERS, Patricia J. (2011), *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*, San Francisco: Jossey Bass/Wiley.
- GERTLER, Paul J., e al. (2011), *Impact Evaluation in Practice*, The World Bank.
- GLASS, Gene V. (1997), Interrupted Time Series Quasi-Experiments, in Jaeger, Richard M. *Complementary Methods for Research in Education*, Washington, D.C., American Educational Research Association.
- GREEN, Jennifer C. [et al] (1989), Toward a Conceptual Framework for Mixed-Method Evaluation Designs, *Educational Evaluation and Policy Analysis*, Vol. 11, Nº 3; 255-274.
- GUIJT. Irene [et al.] (2011), *Evaluation Revisited: Improving the Quality of Evaluative Practice by Embracing Complexity*, Wageningen UR e Context, international cooperation.

- HUGHES, Karl e HUTCHINGS, Claire (2011), Can we obtain the required rigour without randomisation? Oxfam GB's non-experimental Global Performance Framework, Working Paper, Nº13, International Initiative for Impact Evaluation.
- JERVE, Alf Morten e VILLANGER, Espen (2008), *The Challenge of Assessing Aid Impact: A Review of Norwegian Evaluation Practice*, Evaluation Department, NORAD, Study 1/2008.
- LEEUW, Frans e VAESSEN, Jos (2009), *Impact Evaluations and Development Nonie Guidance on Impact Evaluation*, Network of Networks on Impact Evaluation, Washington D.C.
- MAYNE, John (2008), *Contribution analysis: An approach to exploring cause and effect*, Institutional Learning and Change Initiative, ILAC Brief, nº 16.
- MORH, Lawrence (1995), *Impact Analysis for Program Evaluation*, Sage Publications Inc.
- OWEN, John e ROGERS, Patricia (1999) *Program Evaluation: Forms and Approaches*, Allen & Unwin e Sage Publications (UK).
- PATTON, Michael Quinn (2008), *Utilization-Focused Evaluation*, Sage Publications
- RAVALLION, Martin (2005), *Assessing the Poverty Impact of an Assigned Program*, in Francois Bourguignon and Luiz Pereira Da Silva (eds) *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*. New York: Oxford University Press.
- RAVALLION, Martin (2009), *Should the Randomistas Rule?*, The Economists' Voice. Vol. 6, Nº 2: 1-5.
- ROGERS, Patricia (2008), *Four key tasks in impact assessment of complex interventions*, apresentação realizada no âmbito da conferência *Rethinking Impact: Understanding the Complexity of Poverty and Change*, Cali, Colombia.
- ROGERS, Patricia (2008), *Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions*, Evaluation, Vol. 14 (1): 29-48.
- ROGERS, Patricia J., (2012), *Introduction to impact evaluation*, Impact Evaluation Notes, Nº1, InterAction e The Rockefeller Foundation.
- ROSSI, Peter, LIPSEY, Mark, FREEMAN, Howard (2004) *Evaluation: A Systematic Approach*, Sage Publications.
- SCRIVEN, Michael (2005), *Can We Infer Causation From Cross-Sectional Data*, National Academy of Sciences.
- SCRIVEN, Michael (2006), *Converting Perspective to Practice*, Journal of Multidisciplinary Evaluation, Nº 6: 8-9.
- SCRIVEN, Michael (2008), *A Summative Evaluation of RCT Methodology: & An Alternative Approach to Causal Research*, Journal of Multidisciplinary Evaluation, nº 9: 11-24.
- STERN, Elliot [et al.] (2012), *Broadening the Range of Designs and Methods for Impact Evaluations*, Working Paper, Nº 38, DFID.
- VAESSEN, Jos, (2010) *Challenges in impact evaluation of development interventions: opportunities and limitations for randomized experiments*, Institute of Development Policy and Management e University of Antwerp.
- VAN DEN BERG, Rob D. (2011), *Evaluation in the context of global public goods*, Evaluation, Vol. 17, Nº 4: 405-415.
- VEDUNG, Ervert (2009), *Public Policy and Program Evaluation*, Transaction Publishers.
- WHITE, Howard (2008), *Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation*, NONIE Working Paper, nº7.
- WHITE, Howard (2009), *Some Reflections on Current Debates in Impact Evaluation*, Working Paper, nº 1, International Initiative for Impact Evaluation.
- WHITE, Howard e PHILLIPS, Daniel (2012), *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*, Working Paper, nº 15, International Initiative for Impact Evaluation.

WHITE, Howard, Sinha, Shampa e Flanagan, Ann (2006) *A Review of the State of Impact Evaluation.*