



COOPERAÇÃO  
PORTUGUESA

**IPAD** Instituto Português  
de Apoio ao Desenvolvimento

## **DOCUMENTOS DE TRABALHO**

# **A AVALIAÇÃO PROGRAMA / PAÍS Contributo para debate interno**

Documento de Trabalho Nº 4

Maria João Robalo

GAAI

Outubro 2011

Lisboa, Outubro de 2011

Instituto Português de Apoio ao Desenvolvimento (IPAD)

Av. da Liberdade, 192

1250-052 Lisboa

Os Documentos de Trabalho do IPAD são produzidos por funcionários e colaboradores do IPAD.

Os Documentos de Trabalho são publicados sob responsabilidade única dos seus autores e não reflectem necessariamente a opinião e posição do IPAD.

## 1. Introdução

A avaliação é uma ferramenta poderosa para se perceber/aprender o que resulta, o que não resulta e as razões pelas quais isso acontece. É uma abordagem sistemática acerca do mérito de qualquer “objecto”. O grande objectivo da avaliação é determinar o valor ou o mérito do que se avalia. A avaliação é também uma ferramenta importante em termos de prestação de contas e responsabilização. Por estas razões é necessário que as intervenções da cooperação para o desenvolvimento sejam avaliadas.

Dito isto, levanta-se um conjunto de questões ligado às abordagens, metodologias e instrumentos a utilizar nas avaliações. Assim, a função avaliação deve ser sujeita a debate interno, nas suas componentes de independência, qualidade e utilidade, no sentido de a preparar melhor para os desafios que se vão apresentar, necessariamente, no futuro.

O sistema de avaliação do IPAD baseia-se, fundamentalmente, nos standards internacionais de avaliação aceites na generalidade. A avaliação, queremos acreditar, serve a gestão da organização e das intervenções, fornecendo informação essencial para as suas decisões em termos da relevância, eficácia e desempenho operacional. Estes aspectos ganham uma importância ainda maior num período de recursos escassos, de grande contenção orçamental e necessário rigor nos gastos com os recursos dos contribuintes. Neste contexto, a avaliação, como já referido, tem um papel chave a desempenhar na aprendizagem organizacional, prestação de contas e responsabilização.

Embora acreditando nos aspectos atrás referidos e talvez até por causa disso, achamos também que há um conjunto de acções que podem ser tomadas ao nível da Unidade de Avaliação no sentido de melhorar o seu desempenho. É necessário que a avaliação reforce a qualidade da análise e obtenha resultados, isto é, que seja um meio chave na contribuição para a distinção entre sucesso e fracasso!

Na sequência das últimas avaliações, efectuadas pelo GAAI, nomeadamente as avaliações dos Programas Indicativos de Cooperação (PIC) com Angola, Guiné-Bissau e Timor-Leste, surgiu claramente a necessidade de aprofundar o que são e podem ser este tipo de avaliações.

O objectivo deste documento de trabalho é contribuir para o debate sobre a problemática da avaliação e, neste caso, mais centrado na avaliação de Programa/Estratégia país (PIC), sua utilidade versus complexidade, com os correspondentes custos associados.

Neste âmbito, discute-se a necessidade de melhorar os métodos para identificar os impactos da ajuda ao desenvolvimento, isto é, como podem os avaliadores estabelecer relações de causa/efeito nas suas análises e fornecer recomendações para melhorar a qualidade das avaliações de impacto da ajuda. Discute-se também a complexidade associada a este tipo de avaliações e o seu custo/benefício.

O GAAI tem, desde logo, a responsabilidade pela elaboração dos Termos de Referência (TdR) tanto para as avaliações internas, como para as externas. Cada avaliação é composta por várias fases, nomeadamente a fase de análise documental e a fase de deslocação ao terreno, complementadas com entrevistas e finalmente a produção do relatório final. A primeira fase, relativa à análise documental, deve servir também para refinar os TdR e para a escolha da metodologia mais adequada.

A elaboração dos TdR reveste-se de uma importância capital. É neste âmbito que se discutem os objectivos da avaliação, o que se pretende ver respondido por esta e a metodologia mais adequada para se atingirem esses objectivos.

As metodologias de avaliação vão desde análises qualitativas baseadas em documentos de projecto e entrevistas até análises quantitativas, econométricas ou outras, baseadas em dados quantitativos observados.

A questão central deste documento é levantar um conjunto de questões relacionado com estas metodologias e tentar equacionar como se pode abordar a escolha da metodologia mais adequada, que estará sempre ligada ao objectivo e às questões que se querem ver respondidas. Só para dar um exemplo, os TdR para uma avaliação de impacto, por exemplo, podem e às vezes devem enfatizar a necessidade de uma análise contrafactual credível. Para isso a equipa de avaliação deve integrar pessoas com as devidas competências técnicas para implementar este tipo de métodos.

O presente documento apresenta alguns conceitos relativos ao tipo de avaliações em articulação com as diferentes metodologias de análise. Levanta também, neste âmbito, um conjunto de questões:

- É possível avaliar as estratégias país como programas coerentes?
- O que é que este tipo de avaliações pode oferecer e a quem?
- Que recursos são necessários para as avaliações das estratégias país?
- Como isolar a componente da ajuda de um doador?
- Como estabelecer a causalidade / efeitos das intervenções?
- Como definir a complexidade da análise, tendo em consideração os constrangimentos?

O documento está organizado em 6 pontos. No primeiro, faz-se uma breve análise do que tem sido o Programa Indicativo de Cooperação (PIC); no segundo, abordam-se as metodologias de avaliação; no terceiro, refere-se, sucintamente, o que têm sido globalmente as avaliações dos Programas país (PIC); no quarto, dá-se conta de alguma experiência de outros doadores; segue-se no quinto, uma análise do problema e, finalmente, as conclusões.

## **2. Programa Indicativo de Cooperação (PIC) – Estratégia país**

A cooperação bilateral entre Portugal e os seus principais parceiros tem sido desenvolvida, desde 1999, no quadro dos Programas Indicativos de Cooperação (PIC). Os PIC enquadram estrategicamente a cooperação a desenvolver entre Portugal e cada um dos países parceiros, estabelecendo as linhas de força que vão orientar o contributo português para o desenvolvimento económico e social destes países.

Assim, a ajuda de Portugal está, na quase totalidade, enquadrada no que se convencionou chamar PIC. Contudo, temos dúvidas que estes documentos sejam programas no sentido em que todas as partes que o compõem estão interligadas com uma determinada lógica e cuja intenção é, aquando da sua implementação, a preocupação de um trabalho conjunto e articulado entre as várias áreas de intervenção.

Pelo contrário, a maior parte destes programas são muito gerais e consubstanciam-se num conjunto de projectos e programas sectoriais que aparecem em diferentes momentos temporais e pelas mais diversas razões. No máximo, têm uma motivação comum que é a redução da pobreza.

Em tese, um Programa/Estratégia país identifica os objectivos, as linhas estratégicas de actuação, com as respectivas prioridades, que são o resultado da visão estratégica global, tendo em consideração a estratégia do país parceiro. Este Programa tem em consideração os efeitos de programas anteriores, as avaliações e os recursos disponíveis, consubstanciando-se, depois, em projectos e actividades.

O documento descreve, assim, a estratégia da cooperação portuguesa baseada na identificação das prioridades em relação à intervenção no respectivo país parceiro, indica as modalidades da ajuda e o respectivo nível financeiro para um período de médio prazo. É o instrumento formal mais importante como orientador da cooperação com esse país.

A forma geral e algo vaga, como referido, do documento de Estratégia (PIC) causa desde logo problemas a avaliações futuras e ao acompanhamento, que se pretende a um Programa. Apesar disto, o PIC como uma base para o diálogo entre as partes, tem sido muito útil.

Sintetizando as funções da Estratégia país:

- Fornece informação ao país parceiro e indirectamente a outros doadores;
- Serve de base de diálogo (orientação política);
- É um instrumento orientador para toda a cooperação com um país parceiro;
- Identifica as prioridades do país parceiro, que fundamentam as intervenções;
- Identifica os objectivos globais de desenvolvimento, os indicadores e respectivas metas.

A programação é feita com base nas estratégias, nos planos e nas prioridades de desenvolvimento de cada país parceiro, porque é, desde logo, uma forma de promover mais facilmente a sua apropriação.

Na prática, conclui-se então que os PIC têm sido, embora com evolução positiva, um enunciar algo vago, como já referido, da estratégia a desenvolver com o país parceiro e consubstancia-se em intervenções com diferentes tempos de implementação; diferentes executores e diferentes abordagens. Assim, o que é então avaliar o PIC? É avaliar os projectos? É avaliar os sectores? Com diferentes metodologias? É fazer avaliação de efeitos ou de impacto? É avaliar todo o processo de desenho e implementação do Programa?

### 3. Metodologias de Avaliação

Há diferentes tipos de avaliação dependendo do objecto a ser avaliado e do objectivo da avaliação. A distinção básica mais importante entre tipos de avaliação talvez seja entre avaliação formativa (por exemplo avaliação do processo) e avaliação sumativa (avaliação de efeitos, de impacto ou análises custo/eficácia ou custo/benefício).

A avaliação formativa focaliza-se no objecto a ser avaliado – examina o desenho, a implementação e o financiamento do programa. Aborda o contexto organizacional, isto é, os recursos envolvidos, os procedimentos, os *inputs*, etc.

A avaliação sumativa, em contraste, examina os efeitos ou os *outcomes* de um objecto, isto é, o que acontece no final do financiamento do programa. Aborda a questão de se poder afirmar que aquele objecto causou aquele *outcome* e determina o impacto global do factor causal.

As avaliações de Programa país, de uma forma geral, combinam diferentes métodos tanto de análise quanto de recolha de dados/informação. A primeira fonte de informação tem normalmente que ver, para além da análise dos documentos de projecto, com entrevistas aos detentores de interesse, incluindo no país parceiro, e a observação directa pela equipa de avaliação das intervenções mais relevantes, segundo determinados critérios.

A escolha destes detentores de interesse e a recolha de informação raramente têm subjacente critérios estatísticos, baseando-se sobretudo em conhecimentos pessoais ou procura-se encontrar as pessoas no país parceiro que estão por dentro das questões da cooperação e que tenham conhecimento sobre as intervenções.

No entanto, a eficácia e o impacto da ajuda são cada vez mais questões no debate internacional. As agências de desenvolvimento precisam de demonstrar os efeitos do seu apoio nos beneficiários. Estes aspectos levam-nos a metodologias mais complexas, como por exemplo, às avaliações de impacto.

Como referido, em geral as avaliações combinam vários métodos (regressão; inquéritos; entrevistas semi-estruturadas; *focus-group*). O tratamento da informação resulta do cruzamento de várias fontes e métodos de recolha, o que se convencionou chamar **triangulação**. A ideia é que os diferentes métodos e os diferentes tipos de recolha de informação em relação ao mesmo objecto possam suportar-se uns aos outros e tomem as conclusões mais fiáveis.

Há avaliações com abordagens macroeconómicas baseadas em estatísticas oficiais (pode ser um problema nos países mais frágeis) ou em estatísticas próprias recolhidas para esse fim. O objectivo deste tipo de avaliação é muito ambicioso, principalmente no que se refere a determinar os efeitos e impactos atribuídos a uma intervenção de desenvolvimento.

Quanto mais quantitativa e macro for a avaliação (podendo até exigir o tratamento de modelos matemáticos) mais complexa se torna e logo mais dispendiosa. Quando se toma a decisão do tipo de avaliação a ser efectuado e métodos a utilizar tem que se ter presente que provavelmente há uma taxa marginal decrescente de retorno relativamente aos elevados custos que este tipo de avaliações comporta em relação a uma maior fiabilidade. Nestes casos seria útil, aquando da decisão, ter em conta a quantidade de recursos a alocar à avaliação para se obterem dados razoavelmente fiáveis e o que poderia ser atingido com mais dados e análises mais complexas, implicando maiores custos.

É também muito difícil determinar o que se deve à intervenção e o que se deve a intervenções de outros doadores ou até a intervenções por parte do país parceiro. Há dificuldade em ligar os efeitos observados aos resultados das intervenções e aos objectivos do Programa/Estratégia. Coloca-se, então, a questão da **atribuição**, que veremos mais à frente.

### 3.1. Evolução da temática “avaliação”

A primeira evidência formal de avaliação de Programa aparece por volta de 1897-1898 na área da Educação com Joseph Rice. Esta avaliação consistiu num estudo comparativo de 33000 estudantes do sistema escolar dos Estados Unidos relativamente à capacidade de expressão/pronúncia. Depois só por volta dos anos 30 foi elaborado um estudo por Ralph W. Tyler e Eugene R. Smith, que se revelou um esforço notável para avaliar os *outcomes* de um programa educacional a oito anos (1933-1941).

Durante os anos 50 e 60 a avaliação teve um grande impulso, nomeadamente na área educacional e na abordagem experimental. O trabalho de Campbell e Stanley (1966) deu um enorme ímpeto às abordagens experimental ou quase-experimental para avaliações de Programa. Iniciaram-se análises económicas sofisticadas, utilizando métodos quantitativos. Estes métodos quantitativos só ganharam uma aceitação generalizada já nos anos 70 (Malcolm Parlett e David Hamilton - 1972; Robert Stake - 1974; Eisner - 1971; Egon Guba; Yvonna Lincoln). Entretanto, também por esta altura, começa a dar-se um enfoque no “processo” da ajuda no contexto da avaliação. Na sequência da implementação de métodos quantitativos, desenvolveram-se grandes desentendimentos entre os proponentes das abordagens qualitativas e quantitativas. Os anos 70 ficaram, assim, marcados, nesta área, por estas polémicas.

No início dos anos 80, cresceu a preocupação com a eficácia da ajuda. Achava-se que seria muito útil (maiores benefícios de análise) juntar-se (integrar) os dois tipos de abordagem (qualitativa e quantitativa) numa mesma avaliação (Cook e Reichardt, 1979; Blaine R. Worthen, 1981 e Madey, 1982), utilizando diversos métodos. Worthen e Sanders (1987), definem avaliação como “*the formal determination of the quality, effectiveness, or value of a program, product, project, process, objective, or curriculum*” (p.22).

A polémica entre os avaliadores continuou e, nos anos 90, alguns investigadores concluíram que o “desenho” experimental e quase-experimental rigoroso (preconizado, por exemplo, pelas avaliações de impacto) era irrealista pela complexidade que comportava, consumidor de muito tempo e caro, que deveria ser substituído por



abordagens mais rápidas e económicas (Joseph Valadez e Michael Bamberger, 1994). Nesta década assiste-se, entretanto, à difusão dos critérios de avaliação do CAD. Mais tarde à adopção da abordagem “gestão centrada nos resultados”, com reflexos nas metodologias de avaliação.

Actualmente a discussão situa-se ao nível das metodologias para as avaliações de impacto, ligada, desde logo, a algum desentendimento também em relação ao que se considera impacto. Devido ao foco colocado nos resultados, a avaliação de impacto tornou-se uma parte importante na agenda do desenvolvimento na última década. Estas reflexões ligam-se ainda aos Objectivos de Desenvolvimento do Milénio (ODM).

### 3.2. Avaliação de impacto

No seio da comunidade de avaliação do desenvolvimento há, como referido, diferentes ideias sobre o significado de “impacto”. Este aspecto, como facilmente se compreende, influencia as escolhas das metodologias mais adequadas para uma avaliação deste tipo.

Existem sobretudo duas posições bastante diferentes, uma, onde se inclui o CAD/OCDE, que considera impacto como os efeitos de longo prazo (positivos e negativos, primários e secundários) produzidos, directa ou indirectamente, intencional ou não, por uma intervenção de desenvolvimento. Outra abordagem, defendida por alguns investigadores, nomeadamente Howard White e pelo BM, consideram que avaliação de impacto consiste em “ver”/constatar que diferença o programa exactamente provocou, isto é, até que ponto se pode atribuir a uma intervenção específica (**atribuição**): melhorou a vida das pessoas? salvou vidas? Neste caso a avaliação de impacto é fazer a análise com e sem a intervenção: O que é que aconteceu com o programa (registo factual) comparado com o que teria acontecido na ausência deste (exige, em geral, a análise do **contrafactual** – encontrar um grupo de comparação/grupo de controlo). Defendem que a avaliação de impacto é a análise da eficácia do desenvolvimento e não a análise da eficácia da ajuda.

As avaliações de impacto rigorosas (assim consideradas sobretudo pelos defensores da segunda abordagem referida) são análises que medem as alterações (líquidas) nos *outcomes* no âmbito de um grupo particular de pessoas, que podem ser atribuídas a um programa específico, usando a melhor metodologia disponível, fiável e apropriada à questão da avaliação que está a ser investigada (respondida) e a um contexto específico. Estas avaliações de impacto são as que lidam exactamente com o problema da **atribuição**.

A comparação da situação “com e sem” a intervenção é um desafio, uma vez que não é possível observar como a situação teria sido sem a mesma intervenção. É neste ponto que se colocam metodologias quantitativas de análises estatísticas ou outras e também a questão complexa, referida, da **atribuição** (efeitos produzidos por).

As avaliações de impacto, neste caso, estimam a eficácia (em termos dos efeitos) do programa, comparando *outcomes* daqueles que participaram no programa em comparação com aqueles que não participaram nesse programa. O desafio chave na avaliação de impacto é encontrar um grupo de pessoas que não participou no programa, mas que está muito próximo dos participantes.

A avaliação de impacto está, assim, principalmente relacionada com os resultados finais das intervenções no que diz respeito ao bem-estar das populações/comunidades. É uma ferramenta no meio de uma panóplia (como já referido) de ferramentas/métodos de avaliação.

Há autores que mantêm que a causa/efeito é demonstrada quando é possível chegar e quantificar esta **atribuição**. No caso em que é demasiado complexo demonstrar esta atribuição, então o que é necessário é mostrar o contributo/**contribuição** (no sentido da atribuição parcial) do programa para o efeito global. Se este efeito líquido não puder ser demonstrado, então o que deve ser demonstrado é a proporção de impacto que é atribuível ao programa.

Há quem afirme que só é possível a **atribuição** em programas simples. Em programas complexos é difícil distinguir o que é consequência directa da intervenção da de outros factores (Rogers). Atribuir uma mudança nos *outcomes* a uma intervenção específica é muito difícil devido a uma panóplia de factores envolvidos.

As avaliações de impacto com qualidade requerem métodos mistos, isto é, uma combinação entre métodos quantitativos e qualitativos, que podem ser, por exemplo, análises econométricas baseadas em dados de inquéritos e análises qualitativas baseadas em documentos de projecto.

As avaliações de impacto exigem que o conjunto de avaliadores envolvidos tenham um conjunto de competências que lhes permita fazer análises mais rigorosas e que, assim, possa utilizar não só métodos qualitativos, mas também métodos quantitativos. A avaliação de impacto tem forte ênfase sumativa. É uma avaliação final de programa que tem o tempo suficiente para ter efeitos.

#### **4. Avaliação dos Programas Indicativos de Cooperação (PIC)**

As avaliações dos PIC (internas e externas) desencadeadas pelo GAAI têm sido de uma forma geral qualitativas, realizadas a meio percurso (são quase sempre durante o último ano de vigência do Programa) e do tipo formativo.

Faz-se a análise da documentação; trabalha-se um guião para as entrevistas; fazem-se entrevistas (semi-estruturadas ou não) em Portugal e no país parceiro; lançam-se inquéritos (pouco frequentes) e valida-se toda a informação com alguma observação directa a projectos no terreno, baseados em análises qualitativas. Nestes termos e não menosprezando a análise qualitativa, constata-se haver pouca análise quantitativa dos resultados e grande fragilidade na recolha de dados primários.

As avaliações, mesmo as que vão mais longe, acabam por ser ao processo, à relevância e coerência do Programa. Avaliar o processo tem sido analisar o processo da programação e implementação da estratégia, enquanto componentes individuais do Programa país. Estas componentes são avaliadas separadamente sem haver uma abordagem global desse Programa.

As avaliações dos PIC não têm sido avaliações de impacto, desde logo por limitações temporais, dado que uma avaliação de impacto exige que o Programa esteja finalizado e que tenha decorrido um conjunto de anos que permita a evidência dos efeitos. Tem havido também limitação de recursos e dificuldade em encontrar as valências técnicas necessárias. Este tipo de avaliações é muito caro e complexo, implicando um enorme esforço nem sempre recompensado pelos resultados que se conseguem obter.

Mesmo tendo sido avaliações que utilizaram maioritariamente metodologias de tipo qualitativo, constata-se terem tido alguns resultados, dado que também devido a elas (até que ponto é difícil afirmar) algumas alterações na programação das acções foram tendo lugar. Podemos aqui dar alguns exemplos, tanto em relação a avaliação de projectos e sectores, como em relação a avaliações dos PIC. A avaliação do Hospital Central de S. Tomé e Príncipe, cuja intervenção veio a ser descontinuada, como a avaliação recomendava. A avaliação às Bolsas de estudo financiadas pelo IPAD (em finais dos anos 90 representavam cerca de metade do orçamento do então ICP), que foram alteradas na sua filosofia, como recomendava a avaliação (recomendações também do exame do CAD/OCDE). A avaliação às intervenções no Sector da Educação (Guiné-Bissau) também produziu alterações no seu desenho e implementação. Em relação a avaliações dos PIC, podemos dar como exemplo terem sido descontinuados os PAC (programa anual de cooperação) e ter aumentado a concentração da ajuda.

## 5. Experiência de outros doadores

A maior parte das avaliações de Programa país têm sido efectuadas, ultimamente, pelas organizações multilaterais. A **Comissão Europeia** avalia os seus Programas país de uma forma ambiciosa. A maior parte deste trabalho é feito por consultores independentes. Preocupam-se em saber se os objectivos foram atingidos, se os efeitos pretendidos foram alcançados e se tudo foi feito da forma mais eficiente possível.

Para o **Banco Mundial** (BM) as avaliações de Programa país são um instrumento importante para gerir a actividade do banco. O BM preocupa-se com três aspectos da sua actividade num país: os *outcomes* do programa; os efeitos produzidos pelas intervenções de desenvolvimento nas organizações; e a sustentabilidade dos mesmos.

O **Banco Asiático de Desenvolvimento** preocupa-se com a “confiança”. O objectivo das avaliações é claro: são necessárias para a accountability e a aprendizagem. A Estratégia país é normalmente baseada nos objectivos de desenvolvimento do país beneficiário e as “Estratégias de redução da pobreza” são a base da avaliação.

Para as **Nações Unidas** as avaliações devem focar-se nos *outcomes*, mas não tentar encontrar impactos de longo prazo.

A cooperação dinamarquesa (**DANIDA**) efectuou, em 2006, uma avaliação da sua cooperação com o Uganda nos últimos 20 anos. Posteriormente, em 2007, efectuou várias avaliações de Programa país, nomeadamente, entre outros, com o Gana e o Benim. Estas avaliações olham para a relevância da ajuda dinamarquesa nos contextos da política dinamarquesa, da sua estratégia e da estratégia do país parceiro. Apreciam os instrumentos utilizados na cooperação bilateral, as modalidades de ajuda e o peso relativo da ajuda no desenvolvimento económico e social. Estas avaliações são consideradas complexas no sentido em que integram diversos sectores e diversas formas de ajuda.

A cooperação britânica (**DFID**) é, talvez, a cooperação bilateral que mais investe neste tipo de avaliações. Têm tido, inclusivamente, como objectivo fazer cerca de 5 avaliações destas por ano. Em algumas delas o nível de abordagem corresponde a “light approach”, o que significa que utilizam metodologias pouco complexas, mais qualitativas. O método é baseado na análise de documentação disponível complementado por entrevistas no próprio país e no país parceiro. Por norma não visitam os projectos no terreno. Têm abordagens mais ou menos standardizadas, independentemente do país. Assim, as avaliações levam pouco tempo e envolvem poucos consultores, importando em menores custos.

Estas avaliações são, no entanto, mais do que uma mera colecção de perspectivas dos entrevistados, incluem análises e são suportadas por evidências recolhidas durante a avaliação. Estão focalizadas nos objectivos e estratégia do DFID e avaliam a sua performance em relação a cada país parceiro. O interesse destas avaliações reside na forma como captam as perspectivas externas e são sempre análises feitas por avaliadores independentes.

Há alguns doadores bilaterais, como a Holanda e a Noruega que deixaram de fazer avaliações de Programa país, devido aos custos elevados e dada a sua complexidade em relação à sua utilidade. Conclusão retirada após análises custo/benefício.

A cooperação norueguesa (**NORAD**) encetou um debate acerca da questão de como melhorar os métodos de avaliação de impacto das intervenções de desenvolvimento e com isso melhorar a qualidade da ajuda. Questiona como podem os avaliadores estabelecer a causa/efeito nas suas análises. Esta discussão, aliada ao mérito da avaliação Programa país, tem também lugar na cooperação sueca (**SIDA**), que tem vindo a produzir documentos nesta área.

Os relatórios de avaliação de todas estas cooperações mostram, assim, grande diversidade em termos de abordagens metodológicas. Vão desde análises econométricas baseadas em dados de inquéritos até análises qualitativas baseadas em documentos de projecto. Muitas vezes, as análises são mesmo superficiais devido a constrangimentos de recursos.

## 6. Análise

As avaliações consistem numa abordagem analítica, tanto quanto possível, objectiva e sistemática das actividades da ajuda ao desenvolvimento. Levantam questões fundamentais acerca do desenho das intervenções, dos métodos de implementação, dos objectivos, dos efeitos e da sustentabilidade das mesmas.

As avaliações são usadas para melhorar a ajuda ao desenvolvimento fornecendo *feedback* à Organização responsável pela implementação dessa ajuda, ao país parceiro e a outros detentores de interesse sobre os resultados, problemas e desafios relacionados com as intervenções implementadas. A avaliação é, assim, uma fonte de informação chave para a decisão política sobre os resultados e recursos investidos na ajuda ao desenvolvimento.

Na realidade, a prática do desenvolvimento não é tão simples como alguns querem fazer crer. As agências de desenvolvimento não são entidades homogéneas, mas

“arenas” onde divergem os conceitos de desenvolvimento e onde numerosos detentores de interesse têm múltipla agendas.

Cada detentor de interesse, no seu papel, tem que continuamente negociar e renegociar o seu caminho no processo. Todo o staff das agências, aos mais diferentes níveis; consultores; governos; terreno, têm alguma capacidade de influenciar os *outcomes* das intervenções de desenvolvimento.

O desenvolvimento é, de facto, um processo complexo que envolve trabalho e conhecimento em constante mutação e dentro de um discurso/diálogo muitas vezes contraditório. Daqui, desde logo se infere que a avaliação não tem a tarefa facilitada e que por mais simples que sejam as abordagens e os métodos adoptados, existe sempre alguma complexidade inerente ao processo.

Podemos desde já afirmar que não há uma metodologia padrão (*gold standard*), mesmo para avaliações de Programa país, e que usar métodos qualitativos e/ou quantitativos e os seus pesos relativos depende do contexto, do objectivo e das limitações de cada avaliação.

O mais ambicioso é olhar para uma combinação de efeitos de longo prazo do Programa. Isto é algo de muito difícil, desde logo porque é necessário determinar que mudanças ocorreram em resultado do programa individualmente e não de outros factores, conjugados ou não. A ajuda raramente é a única fonte de financiamento de uma intervenção de desenvolvimento pública ou privada e muitas vezes não é a fonte mais significativa. Mas a ajuda não é só uma questão de financiamento directo à intervenção, ela chega de diferentes formas, tais como: contribuição para o orçamento do país parceiro (recursos financeiros); formação; assistência técnica (recursos humanos – transferência de conhecimento). Esta abordagem, que examina os efeitos totais do programa, requer normalmente avultados recursos.

Outro tipo de abordagem tem que ver com a análise das componentes do processo de cooperação. Isto é, a análise relativa à elaboração e implementação do Programa/Estratégia, enquanto componentes individuais do Programa avaliadas separadamente. Esta abordagem tem a vantagem de poder oferecer experiências úteis a ambas as partes (doador e parceiro) no que diz respeito a um determinado processo analisado, podendo ser generalizado e aplicado a outros processos de implementação da Estratégia.

Uma terceira alternativa, por exemplo, é “olhar” só para alguns aspectos em particular, por exemplo, avaliar alguns pressupostos, a consistência e a relevância do Programa.

Há um conjunto de problemas chave relacionados com as avaliações de Programa país. Primeiro, há as dificuldades em identificar a contribuição individual para o desenvolvimento de um país; segundo, há problemas relacionados com os aspectos numéricos dos dados e com a maior ou menor subjectividade; terceiro, há questões relacionadas com a utilidade destas avaliações, isto é, para quem elas são verdadeiramente úteis.

Mesmo quando um sector ou grande projecto (ex. "Saúde para Todos") é dominado por um doador, é muitas vezes impossível ter a certeza das razões de uma mudança observada. Muitos factores externos incontroláveis afectam os acontecimentos. Aqui coloca-se, como já vimos, um problema relacionado com a avaliação de impacto, que é a questão da **atribuição**. Um dos principais desafios na avaliação de impacto é esclarecer o quanto da variação observada pode ser atribuído à intervenção em si.

Outra questão também muito importante e crítica, neste âmbito, é o *timing*. Se as avaliações de impacto são feitas muito cedo, podem não ser observados efeitos importantes. Se as avaliações são feitas muito tarde, pode já não ser fácil avaliar os impactos.

A solução mais comum, embora difícil, para o problema da **atribuição**, acima referida, é a selecção de um grupo de comparação (grupo de controlo). A diferença nos resultados entre os beneficiários da intervenção (grupo de tratamento) e o grupo de comparação pode ser uma indicação dos efeitos a imputar à intervenção. Aqui também se podem levantar problemas, por exemplo, relacionados com a questão de o grupo de tratamento e o grupo de controlo possuírem características diferentes.

Este processo de selecção significa que os beneficiários não são uma amostra aleatória da população, o que significa que o grupo de comparação também não deve ser uma amostra aleatória da população como um todo, mas sim extraídos de uma população com as mesmas características que os escolhidos para beneficiários da intervenção.

Relativamente ao acompanhamento, sendo eficaz é uma ferramenta muito importante para a avaliação. Neste âmbito constata-se, às vezes, uma contradição em relação ao trabalho/empenho que é colocado para elaborar um Programa de ajuda país coerente e o pouco interesse evidenciado pelo acompanhamento deste, enquanto Programa.

Há duas razões fundamentais para a falta de acompanhamento sistemático e ambicioso das Estratégias país. A principal está relacionada com o facto de tanto os doadores quanto os beneficiários se focalizarem nos seus interesses ligados a projectos específicos, em vez de darem atenção à estratégia no seu todo. Os projectos

estão mais perto da actividade operacional e requerem uma supervisão frequente. A outra razão prende-se com o facto de ser metodologicamente muito difícil (tanto para o acompanhamento, como para a avaliação) e eventualmente muito caro identificar os efeitos da cooperação a um nível macro. Os relatórios centram, regularmente, a sua análise na relevância e nos resultados (*outputs*). A sustentabilidade de longo prazo e a eficiência são pouco ou mesmo nada discutidas/abordadas.

Os objectivos e respectivos indicadores são muitas vezes pouco claros na sua formulação. As avaliações têm muitas vezes um carácter de “revisão a meio percurso” e não tanto um carácter de avaliação final (*ex-post*).

As avaliações finais podem focalizar-se nos *outcomes* de médio prazo e não tentar encontrar os impactos que são necessariamente de mais longo prazo. Encontrar ligação directa entre a intervenção e os efeitos detectados é difícil e, naturalmente, que esta dificuldade aumenta com prazos mais longos. O ser-se menos ambicioso na análise não quer, necessariamente, dizer que se é menos rigoroso, mas tão só que se está a colocar o foco nos efeitos de curto e médio prazo, por aumentar a fiabilidade da análise. Acresce, como já referido, que, neste caso, a avaliação necessita de tempos menores e, como consequência, de custos menores.

Há uma grande variedade de tipo de avaliações no que diz respeito à forma e abordagem. Há quem ache que uma definição estrita de “*country programme evaluation*” não é possível.

As motivações podem ser as mais variadas. Podem ser efectuadas análises à *performance* dos programas como um todo, ou comparar sectores e métodos de implementação dentro (no contexto) do programa. Pode-se querer explicar os resultados de todo o programa em relação, por exemplo, às condições políticas ou ao foco sobre o objectivo da redução da pobreza. Pode ser só para fornecer recomendações tendo em vista melhorar os resultados.

Em síntese, coloca-se um conjunto de problemas chave relativo às avaliações de Programa país:

- Dificuldades de atribuição a uma organização específica (a uma agência de desenvolvimento) os seus contributos para o desenvolvimento (contributo individual). Mesmo quando um sector/um grande projecto é dominado por um doador (ex. STP – Saúde), a maior parte das vezes não é possível estar certo das razões para uma mudança observada. Muitos factores externos não controláveis afectam os acontecimentos;



- Problemas metodológicos acrescidos com estas avaliações. Estão desde logo ligados a métodos quantitativos e ao apuramento de dados e a abordagens mais ou menos subjectivas;
- Questões que se colocam ao nível da utilidade/custos destas avaliações;
- Informação produzida pelo acompanhamento insuficiente ou mesmo desadequada para os objectivos da avaliação;
- Problemas na análise dos programas relacionados com a atenção que o país parceiro deveria dar aos mesmos, nomeadamente ao nível da gestão e da alocação de recursos. Uma das formas que pode ajudar a ultrapassar este aspecto bastante problemático e que tem colocado problemas mais ou menos graves relativamente à sustentabilidade das intervenções, são as avaliações conjuntas, onde (neste contexto) o doador e o país parceiro cooperem no sentido de apurar as fragilidades e constrangimentos, sensibilizando-se, num contexto de aprendizagem, o país parceiro para as mesmas;
- Limitações de dados. No contexto do acompanhamento e da avaliação colocam-se muitas vezes problemas com a ausência de **baseline**. Esta ausência de *baseline* pode ser ultrapassada pela recolha de “*secondary data*” (por estimativas que vão implicar a triangulação – comparar estes com resultados de outras abordagens, por exemplo, observação directa; outras fontes secundárias; informadores chave; inquéritos aos detentores de interesse; abordagens participativas; fotografias; artigos de jornal);
- Limitações de tempo;
- Limitações de financiamento;

A equipa de avaliação pode ter que equacionar alguns *trade-off* entre as limitações práticas e objectivas e os requisitos científicos para a avaliação em causa.

Assim, em relação às técnicas de avaliação, está disponível uma grande variedade de ferramentas, diferindo em termos da sua racionalidade, complexidade, dados requeridos e pressupostos. A escolha do método apropriado para um estudo de avaliação específico é uma das dificuldades na elaboração dos TdR. Situa-se aqui o debate, tão necessário até para que, como refere Scriven (“*The Three Revolutions*”), a área da avaliação passe de “intocável” a “respeitável” (primeira revolução).

## 7. Conclusões

A avaliação tem como objectivos principais a aprendizagem e a *accountability*. As avaliações são usadas para melhorar a ajuda ao desenvolvimento, fornecendo “feedback” acerca dos resultados, problemas e desafios relacionados com as actividades em curso. As avaliações são também uma fonte privilegiada de informação para a decisão política.

Segundo Scriven, no documento *The Logic of Evaluation* (2007), “*evaluation is the process of determining merit, worth, or significance; an evaluation is a product of that process. Professional evaluation is evaluation done in a systematic and objective way with degree of expertise that requires extensive specific training or learning*”. A análise tem-se centrado, sobretudo, na eficácia da implementação do programa, em saber se o programa como um todo é/foi relevante e consistente e se foram produzidos, pelo programa, efeitos de médio e longo prazo.

Há fundamentalmente dois tipos de avaliação, uma formativa ligada à implementação do processo nas avaliações a meio percurso, tendo como resultado a aprendizagem no sentido de ajudar a melhorar a “performance” da intervenção. Outra sumativa que consiste na avaliação do sucesso da intervenção depois desta ter terminado. Este tipo de avaliação centra-se nos *outcomes* de médio e longo prazo. Temos, por exemplo, neste caso as avaliações de impacto (este tipo de avaliação é, por exemplo, o mais adequado para analisar a relação causal entre as intervenções e as alterações de comportamento/desempenho).

Uma das principais preocupações é, efectivamente, ver respondidas as questões da avaliação e para isso a selecção da metodologia/métodos de avaliação é muito importante. É a maneira pela qual (método/s utilizado) a análise vai ser abordada. Esta questão deve estar, desde logo, explícita nos TdR da avaliação. Daqui a necessidade de reflexão em relação ao que se quer avaliar, que respostas se querem obter, que constrangimentos se colocam, que tipo de avaliação estamos em condições de fazer, com que metodologias, que sejam as mais adequadas, no sentido de se atingir o objectivo da avaliação.

Os TdR necessitam de ser consistentes, com grande coerência interna, reflectindo a abordagem metodológica realista possível, mais ou menos científica, tendo em conta os dados, o tempo e o orçamento disponíveis. O *timing* da avaliação é crítico em termos de abordagem metodológica. Tem que ver com o tempo necessário para que os efeitos e impactos surjam.

Ao optar-se, por exemplo, por avaliações de impacto, tem que se ter presente que este tipo de avaliação pode significar custos mais elevados e pode exigir conhecimentos específicos acerca de metodologias algo complexas de avaliação. Pode mesmo exigir competências de análise avançadas.

Neste âmbito, podemos afirmar que:

- Nenhum método em particular é melhor para tratar o conjunto de questões das avaliações de impacto;
- Contudo, dependendo das questões específicas ou dos objectivos de uma dada avaliação de impacto, alguns métodos têm vantagem em relação a outros na análise dessas questões;
- Diferentes métodos podem complementar-se para dar uma análise de impacto mais completa.

Processos/métodos puros de avaliação são muito raros. A maior parte das avaliações, mesmo as de impacto, utilizam métodos quantitativos e qualitativos. As boas avaliações são, quase sempre, avaliações com método misto, que inclui informação e análises qualitativas e quantitativas.

Utilizar um conjunto de métodos ou informação proveniente de diferentes abordagens (triangulação) pode ser a forma mais eficaz de se obter bons resultados com as avaliações. Por exemplo, pode-se numa única avaliação utilizar uma conjugação de entre vários métodos, por exemplo, método experimental – *RCT (randomized control trials)*; “*theory-based approaches*”; inquérito; estudo de caso; entrevistas semi-estruturadas; *focus group*. No fundo o que se pretende é estruturar a informação no sentido de se conseguir contar uma “história clara”.

Assim, facilmente se compreende que a reflexão acerca da escolha das metodologias (mais qualitativas ou mais quantitativas; análises de processo ou de impacto; etc.) de avaliação, no âmbito da função avaliação, seja tão importante.

Esta reflexão é complexa, exige, desde logo, vontade, mas também conhecimentos técnicos sobre estas temáticas. Se nos focalizarmos exactamente nas avaliações de Estratégia país (PIC), estamos confrontados com Programas, cujas intervenções têm executores diferentes, abordagens de implementação diferentes e *timings* diferentes.

Talvez seja por estes aspectos, que tornam complexa a avaliação do Programa como um todo, que várias cooperações bilaterais não se têm focalizado muito na avaliação de Programa país. Há quem (por exemplo, a cooperação sueca - SIDA) ultrapasse estes aspectos, propondo avaliações autónomas das intervenções do Programa.

Nesta matéria há quem advogue as avaliações conjuntas entre doadores para colmatar alguma da complexidade, desde logo, aliada à **atribuição**, deste tipo de avaliações. Achar mesmo paradoxal que se fale e se focalize neste tipo de avaliações individuais (cada doador) de Programa país, quando a Declaração de Paris sobre a eficácia da ajuda se foca na coordenação e harmonização de métodos de trabalho.

Tanto quanto possível as avaliações devem basear-se no material existente, levar pouco tempo e envolver poucos recursos materiais e humanos. No entanto, tem que se ter em atenção a qualidade da informação/dados empíricos. Há um limite para uma abordagem “*light*” porque coloca em causa a fiabilidade da avaliação. Muitas vezes um método um pouco mais rigoroso não implica necessariamente mais dados.

É muito comum restringir a recolha de dados às entrevistas, com informadores chave e grupos focais. Aproveitar a experiência de campo dos avaliadores pode colmatar algumas insuficiências, dado ser uma fonte inestimável de dados qualitativos que não deve ser menosprezada.

Bons dados de base são essenciais para qualquer avaliação. Quando se trabalha com dados secundários, a ausência de informação em relação à qualidade do processo de recolha destes dados pode restringir as opções de análise dos dados e colocar em causa os seus resultados.

As avaliações são muitas vezes, como já referido, mais exames/análises a meio percurso do que avaliações macro finais. Nestes casos pode usar-se um peso maior de metodologias qualitativas. Ainda assim convém não exagerar na utilização única de métodos qualitativos. Deve-se tentar balancear a análise qualitativa com a quantitativa.

Deve evitar-se, a todo o custo, desenhos desadequados, dados inválidos, análises demasiado superficiais e relatórios irrelevantes em relação às questões importantes que deveriam ser colocadas e respondidas pela avaliação.

A arte de uma boa avaliação está na elaboração das perguntas certas e na recolha e análise adequada dos dados para poder responder a essas questões num contexto específico.

## Bibliografia

- American Evaluation Association (AEA), 2009, <http://www.eval.org/>
- Bamberger, Michael; Rugh, J. and Mabry, L., 2006, RealWorld evaluation: Working under budget, time, data and political constraints.
- Barahona, Carlos, 2010, Randomised Control Trials for Impact Evaluation of Development Initiatives: A Statistician's Point of View. [http://www.biversityinternational.org/fileadmin/biversity/publications/pdfs/1391\\_Randomised%20control%20trials%20for%20the%20impact%20evaluation%20of%20development%20initiatives.pdf](http://www.biversityinternational.org/fileadmin/biversity/publications/pdfs/1391_Randomised%20control%20trials%20for%20the%20impact%20evaluation%20of%20development%20initiatives.pdf)
- Blamey, Avril; Judge, Ken and Mackenzie, Mhairi, 2002., Theory-Based Evaluation of Complex Community-Based Health Initiatives. [http://www.nice.org.uk/niceMedia/pdf/SemRef\\_TheoryBased\\_Blamey.pdf](http://www.nice.org.uk/niceMedia/pdf/SemRef_TheoryBased_Blamey.pdf)
- Campbell, D. T., and Stanley, J. C., 1966, Experimental and Quasi-experimental Designs for Research.
- Chambers, Robert; Karlan, Dean; Ravallion Martin and Rogers, Patricia, 2009, Designing impact evaluations: different perspectives. [http://www.3ieimpact.org/admin/pdfs\\_papers/50.pdf](http://www.3ieimpact.org/admin/pdfs_papers/50.pdf)
- Cook, T. D. and Reichardt, C. R., 1979, Qualitative and Quantitative Methods in Evaluation Research.
- European Centre for Development Management (ECDPM), s.d, Capacity, Change and Performance, [http://www.ecdpm.org/Web\\_ECDPM/Web/Content/Navigation.nsf/index2?readform&http://www.ecdpm.org/Web\\_ECDPM/Web/Content/Content.nsf/0/77BA29B424499B59C12573FB0054BEDB?OpenDocument](http://www.ecdpm.org/Web_ECDPM/Web/Content/Navigation.nsf/index2?readform&http://www.ecdpm.org/Web_ECDPM/Web/Content/Content.nsf/0/77BA29B424499B59C12573FB0054BEDB?OpenDocument)
- Evaluation Cooperation Group (ECG), 2008, Good Practice Standards for Country Strategy and Program Evaluations, [www.ecgnet.org/gps/country\\_strategy\\_program](http://www.ecgnet.org/gps/country_strategy_program)
- Green, Beth L. and McAllister, Carol, 1998, Theory-Based, Participatory Evaluation: A Powerful Tool for Evaluating Family Support Programs. <http://www.npcresearch.com/Files/SRCDZ3.pdf>
- Independent Evaluation Group (IEG) / World Bank – Conducting quality impact evaluations under budget, time and data constraints, 2006, <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20648877~isCURL:Y~menuPK:435292~pagePK:148956~piPK:216618~theSitePK:384329,00.html>
- Independent Evaluation Group (IEG) / World Bank, 2011, Writing terms of reference for an evaluation: A How-To Guide, [http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/ecd\\_writing\\_TORs.pdf](http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/ecd_writing_TORs.pdf)
- International Initiative for Impact Evaluation (3ie–Impact Evaluation, Development Effectiveness), s.d., Principles for Impact Evaluations, <http://www.3ieimpact.org/doc/principles%20for%20impact%20evaluation.pdf>
- Lundgren, Hans E.; North, W. Haven and Rist, Ray C., 2003, A Review of Evaluation in Danida.
- McAlister, Alfred L., 1991, Population Behaviour Change: A Theory-Based Approach, <http://www.bvsde.paho.org/bvsacd/cd68/paho557/s19.pdf>
- Network of Networks on Impact Evaluation – Impact Evaluations and Development – NONIE Guidance on Impact Evaluation, 2009, [http://siteresources.worldbank.org/EXTOED/Resources/nonie\\_guidance.pdf](http://siteresources.worldbank.org/EXTOED/Resources/nonie_guidance.pdf)
- Norad, 2008, The Challenge of Assessing Aid Impact: A review of Norwegian Evaluation Practice, Study 1/2008, <http://www.norad.no/en/Tools-and-publications/Publications/Publication-Page?key=109790>
- Owen, John M. and Rogers, Patricia J., 2006, Program Evaluation. Forms and Approaches.
- Sasaki, Ryoh, 2006, A Review of the History and the Current Practice of Aid Evaluation, [http://survey.ate.wmich.edu/jmde/index.php/jmde\\_1/article/view/51/58](http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/51/58)
- Scriven, Michael, 2007, The Logic of Evaluation.
- Sida – Evaluations Newsletters - <http://www.sida.se/Svenska/Om-oss/Sa-arbetar-vi/Utvardering/Sida-Evaluation-Newsletter/Sida-Evaluations-Newsletter-2009-2/>
- Sida, 2007, Evaluations of Country Strategies, <http://sidapublications.citat.se/interface/frmoptimaker3.asp?doctype=3&order=createdate%20DESC&departmentid=298&topheight=55&headerheight=23&fotheight=0&leftframe>

[width=300&width=820&stylesheet=sida.css&frameout=0&language=14&login=True&username=sida2&password=sida2](#)

- Vaessen, Jos, 2010, Challenges in impact evaluation of development interventions: Opportunities and limitations for randomized experiments
- Valadez, Joseph and Bamberger, Michael, 1994, Monitoring and Evaluating Social Programs in Developing Countries. [http://www.pol.ulaval.ca/perfeval/upload/publication\\_192.pdf](http://www.pol.ulaval.ca/perfeval/upload/publication_192.pdf)
- Van Den Berg, Rob D, 2011, Evaluation in the context of global public goods, [http://www.thegef.org/gef/sites/thegef.org/files/documents/Evaluation\\_PublicGoods\\_Co\\_next.pdf](http://www.thegef.org/gef/sites/thegef.org/files/documents/Evaluation_PublicGoods_Co_next.pdf)
- White, Howard, 2009, Some Reflections on Current Debates in Impact Evaluation, [http://www.3ieimpact.org/admin/pdfs\\_papers/10.pdf](http://www.3ieimpact.org/admin/pdfs_papers/10.pdf)
- White, Howard, 2009, Theory-Based Impact Evaluation: Principles and Practice, [http://www.3ieimpact.org/admin/pdfs\\_papers/48.pdf](http://www.3ieimpact.org/admin/pdfs_papers/48.pdf)
- White, Howard, 2011, An Introduction to the use of randomized control trials to evaluate development interventions, [http://www.3ieimpact.org/admin/pdfs\\_papers/Working%20Paper%209%20Final.pdf](http://www.3ieimpact.org/admin/pdfs_papers/Working%20Paper%209%20Final.pdf)
- Worthen, B. R and Sanders, J. R., 1987, Educational Evaluation: Alternative Approaches and Practical Guidelines.
- Worthen, B. R., 1990, Program Evaluation. <http://legacy.oise.utoronto.ca/research/field-centres/ross/ct1014/Worthen1990.pdf>