

Modelo de Avaliação de Políticas Públicas – MAPP-CIES. Relatório Final.

Janeiro de 2023

Programa Operacional Assistência Técnica (POAT)

Ficha técnica

Título: Modelo de Avaliação de Políticas Públicas – MAPP-CIES. Relatório Final.

Coordenadores: Luís Capucha e João Sebastião

Autores: Alexandre Calado, Ana Rita Capucha, Elisa Alves, Eva Gonçalves, João Henriques, Nuno Nunes, Pedro Estêvão.

Data: Janeiro de 2023

Financiamento: Programa Operacional Assistência Técnica (POAT)



Índice

RESUMO / ABSTRACT	1
INTRODUÇÃO	2
ENQUADRAMENTO E CONTEXTO DO OBJETO DE ESTUDO	3
Descrição do problema	5
Referenciais Teóricos	7
Estratégia metodológica	18
PRINCIPAIS CONCLUSÕES	21
REFERÊNCIAS BIBLIOGRÁFICAS	23
ANEXOS	25
Anexo A: Grelha de análise aos relatórios de avaliação	25
Anexo B: Grelha de análise da Avaliação do Contributo do PT2020 para o POCH	27
Anexo C: Grelha de análise da Avaliação de Impacto dos CLDS	32
Anexo D: Guião da entrevista (Coordenadores e Consultores)	37
Anexo D: Relatórios/transcrições das entrevistas	40

Resumo / Abstract

O estudo Modelo de Avaliação de Políticas Públicas do CIES-IUL (MAPP-CIES) visou o desenvolvimento de uma abordagem teórico-metodológica à avaliação de políticas públicas baseada na Teoria da Mudança, que utiliza, entre outros, o método contrafactual para a avaliação de impactos o seguindo uma perspectiva compreensiva e uma estratégia “multimétodo”. A estratégia metodológica baseada na análise aprofundada da literatura científica sobre a Teoria da Mudança e a Avaliação de Impacto Contrafactual; a análise de dois casos de estudo de avaliação de políticas públicas (Avaliação de Impacto dos Contratos Locais de Desenvolvimento Social e a Avaliação do Contributo do PT2020 para a Promoção do Sucesso Educativo, Redução do Abandono Escolar Precoce e Empregabilidade dos jovens para o PO CH); e entrevistas com stakeholders institucionais pertinentes e peritos de avaliação em políticas públicas. As conclusões apontam para um modelo de avaliação baseado numa abordagem da TdM que acompanha todo o ciclo de vida da política, nomeadamente a identificação do problema, o planeamento, a implementação e os resultados e impactos previstos, sendo assim operacionalizada em três níveis: i) teoria programática; (ii) teoria da operacionalização; (iii) teoria da produção de efeitos. Dadas as limitações inerentes ao método contrafactual, a sua prescrição em matéria de modelos de avaliação deve ser criteriosa e, quando aplicada, deve ser articulada com a TdM, que lhe providencia uma orientação substantiva. Os resultados do MAPP-CIES foram integrados no Guião de Orientação MAPP-CIES, orientado para futuros avaliadores de políticas públicas, que inclui, para cada dimensão de análise, as questões a responder, a estratégia metodológica a seguir e os modos de contornar alguns dos constrangimentos que geralmente limitam, por um lado, o acompanhamento dos programas e, por outro lado, os estudos de impacto.

The CIES-IUL Public Policy Evaluation Model study (Mapp-CIES) aimed at the development of a theoretical-methodological approach to public policy assessment based on the Theory of Change (TdM), which uses, among others, the counterfactual method (AIC) for the evaluation of Impacts. This model follows both a comprehensive perspective and a “multimethod” strategy to evaluation. The methodological strategy was made based on: the in-depth analysis of the scientific literature on the theory of change and the counterfactual impact assessment; the analysis of two cases of public policy evaluation study [Impact Evaluation of Local Social Development Contracts (CLDS) and the Evaluation of the PT2020 Contribution to the Promotion of Educational Success, Reduced Early School Abandonment and Employability of Young People (PO CH)]; and interviews with institutional stakeholders and policy evaluation experts. Research findings point to an evaluation model based on an TdM approach that accompanies the entire intervention life cycle, namely the conception, planning, implementation and planned results and impact phases, thus being operationalized at three levels: I) programmatic theory; (ii) implementation theory; (iii) theory of effect production. Given the limitations inherent to the AIC, its prescription in the matter of evaluation models must be careful and, when applied, must be articulated with the TfM, which provides a substantive orientation. The results of Mapp-CIES were integrated into the Mapp-CIES Orientation Guide to future public policy evaluators, which includes, for each phase of analysis, the evaluation questions, the following methodological strategy and the ways of circumventing some of the constraints that usually limit, on the one hand, the monitoring of the programs and, on the other hand, impact studies.

Introdução

O presente relatório apresenta um resumo dos trabalhos realizados e dos principais resultados do Estudo Modelo de Avaliação de Políticas Públicas do CIES-IUL (MAPP-CIES), no âmbito do concurso para Estudos sobre Metodologias de Monitorização e Avaliação, Eixo I-Coordenação, Gestão Monitorização e Auditoria, Aviso N.º 06/POAT/2021. Para o Estudo foram definidos dois principais objetivos:

1. Desenvolver um Modelo de Avaliação de Políticas Públicas (MAPP-CIES) elaborado numa perspetiva compreensiva, baseada na Teoria da Mudança e assente numa estratégia metodológica “multimétodo”, que utiliza, entre outros, o método contrafactual para a avaliação de impactos, e testar esse modelo a partir de dois estudos previamente realizados;
2. Criar um Guião de Orientação MAPP-CIES para futuros avaliadores de políticas públicas (cofinanciadas pelos fundos europeus e outras) que inclua, para cada dimensão de análise, as questões a responder, a estratégia metodológica a seguir e os modos de contornar alguns dos constrangimentos que geralmente limitam, por um lado, o acompanhamento dos programas e, por outro lado, os estudos de impacto.

O propósito desta investigação passa, assim, por apresentar uma metodologia de avaliação de políticas públicas que consideramos: (i) inovadora, articulando a Teoria da Mudança com a técnica de análise de Avaliação de Impacto Contrafactual; (ii) aplicável a qualquer tipo de intervenção, independentemente da área em estudo (mas que requer o domínio dos temas substantivos para a (re)construção da Teoria da Mudança), e (iii) sensível aos contextos sociais, culturais e económicos, onde sejam implementadas as intervenções de forma a melhorá-las no espírito das prioridades definidas no âmbito da Política da Coesão em Portugal.

Para alcançar os objetivos propostos a equipa de projeto definiu uma metodologia de trabalho flexível e versátil que assenta em três aspetos fundamentais: 1) análise aprofundada da literatura científica sobre a Teoria da Mudança (TdM) e a Avaliação de Impacto Contrafactual (AIC); 2) análise aprofundada de dois casos de estudo de avaliação de políticas públicas, a Avaliação de Impacto dos Contratos Locais de Desenvolvimento Social (CLDS) e a Avaliação do Contributo do PT2020 para a Promoção do Sucesso Educativo, Redução do Abandono Escolar Precoce e Empregabilidade dos jovens para o PO CH (PO CH); e 3) recolha de informação junto de stakeholders institucionais pertinentes, e de especialistas em metodologias de avaliação em políticas públicas, através de um conjunto de entrevistas e focus group.

Do ponto de vista da sua estrutura, este relatório é composto por três seções principais: (i) Enquadramento e contexto do objeto de estudo, (ii) apresentação dos contributos do estudo, e (iii) discussão das principais conclusões. O enquadramento e contexto do objeto de estudo discute os fundamentos e o estado da avaliação de políticas públicas. Nos contributos mais relevantes do estudo, em primeiro lugar é apresentada uma descrição do problema, tendo como referência a discussão em torno da TdM e da AIC, e onde se apresentam as principais conclusões das entrevistas realizadas. Em segundo lugar são apresentados os principais produtos dos referenciais teóricos da TdM e da AIC, E em terceiro lugar é explanada a estratégia metodológica adotada. O relatório termina com uma síntese das principais conclusões.

Enquadramento e contexto do objeto de estudo

A avaliação faz parte do ciclo de vida de uma política pública. No âmbito da governação pública, a avaliação cumpre com os princípios de transparência democrática, mas também da eficiência gestonária do policy-making público, destacando-se neste ponto, ainda, o domínio da prestação de contas, ou como é designado na literatura especializada, da “accountability”.

A avaliação tem, assim, um papel fundamental, quer na implementação de uma cultura de responsabilidade e de monitorização, na forma de fazer e pensar a política pública (Rodrigues, 2011), quer na promoção de uma cultura de ampla participação, contribuindo para sociedades mais informadas e esclarecidas em torno de matérias de interesse coletivo (Sebastião, Capucha, Martins e Capucha, 2019).

Noutra perspetiva, a avaliação permite o enquadramento de contextos de melhoria continuada, incentivando a transformação e modernização das estruturas administrativas públicas, dos recursos, e empreendendo um maior rigor na delimitação dos objetivos e das metas que assistem ao desenho e implementação de medidas que visam responder a problemas específicos – uns de alcance mais coletivo, outros de alcance mais circunscrito.

Em todo o caso, e apesar de haver já um campo técnico-científico de avaliação no país, ele encontra-se ainda pouco institucionalizado na prática política mais comum, existindo, por isso, espaço para melhorias, sobretudo considerando os aspetos mais reflexivos que acompanham estes processos e que podem devolver importantes pistas para a consolidação de práticas e procedimentos de avaliação das políticas públicas em Portugal, nos planos nacional, regional e local, mais sofisticados e de âmbito mais sistemático. Da mesma forma, os sistemas de monitorização e de acompanhamento previstos nos normativos dos programas e políticas públicas são função de pesados aparelhos administrativos e burocráticos, sendo, por isso, mais limitados para responder à capitalização das boas práticas e à inovação, que agora tanto se valoriza na gestão das políticas públicas (Nascimento e Capucha, 2017).

A avaliação faz parte de um ciclo de uma política pública, porque esta envolve investimento e recursos públicos, sobre os quais é preciso prestar contas; porque envolve uma grande diversidade de atores e de stakeholders; e porque em termos da filosofia política e da ética, se deve pautar por critérios de justiça e equidade social e territorial. Isto obriga, de facto, a construir uma metodologia de avaliação suficientemente robusta, capaz de responder a diferentes objetivos e a captar a complexidade que reveste a implementação das políticas, em especial as que são objeto de financiamento europeu que têm associados agendas, metas e objetivos convergentes com uma série de outros países.

Um modelo de avaliação compreensivo, como aquele que propomos com a realização do MAPP-CIES, capaz de responder a diferentes cenários de avaliação de políticas públicas, envolvendo diferentes metodologias (tanto qualitativas, como quantitativas), que esteja ancorado na construção de uma Teoria da Mudança, oferece inúmeras potencialidades: (1) possibilita um mapeamento teórico das hipóteses que apontam para os fatores, e ou, mecanismos impulsionadores de mudança, sobretudo a identificação das variáveis alteráveis capazes de explicarem a alteração/estagnação das condições de partida; (2) introduz uma lógica processual no tratamento da mudança, que resulta da cadeia de relações causais entre os diferentes fatores operantes nas medidas em avaliação; (3) facilita a abordagem ao terreno e a condução do trabalho empírico, deixando mais claros os domínios que devem ser devidamente explorados; (4) permite que o programa e ou medida em avaliação, seja objeto de melhorias ao longo do programa de avaliação, contribuindo adicionalmente para a aprendizagem e para cenários

futuros de elaboração de políticas mais adaptadas às necessidades e à realidade (Weiss, 1995). Não obstante as potencialidades, as advertências são também algumas, sobretudo porque a construção de uma Teoria da Mudança exige um trabalho periférico reflexivo sobre a própria avaliação em curso, o que não é habitualmente mobilizado nestes processos, e porque subsiste uma tendência na prática avaliativa para tornar mais importantes os resultados, quando a Teoria da Mudança mostra, justamente, a relevância dos resultados intermédios e a importância das fases iniciais dos programas e medidas em avaliação. Além disso, frequentemente, a complexidade das teorias e a falta de ferramentas dos atores para as dominarem, podem dificultar a participação e o envolvimento de todos os atores relevantes nas tarefas de avaliação.

A Teoria da Mudança, neste cenário, surge ancorada a um determinado contexto de onde saem sinalizados os problemas que justificam uma intervenção política. Surge, por isso, como a melhor metodologia a explorar no processo de teste e de melhoria do MAPP-CIES, que assenta no pressuposto de que as políticas públicas são processos de mudança com diferentes fases. Iniciam no momento da sua conceção, vão sendo modificados durante o seu planeamento e a sua concretização, produzindo certos resultados que foram sendo determinados pela forma como os processos foram concebidos, planeados e concretizados e, também, participados pelos vários agentes e atores nas quatro fases que constituem as políticas públicas.

O desenvolvimento do modelo de avaliação de Políticas Públicas (MAPP-CIES) assentou na articulação de duas peças essenciais: (1) a abordagem da Teoria da Mudança, de acordo com uma interpretação compreensiva e cumulativa face às fases que compreende, incorporando ainda o modelo proposto por Stufflebeam (1983) CIPP – contexto, inputs, processos, produtos); (2) a abordagem da Avaliação de Impacto Contrafactual, que interage com a Teoria da Mudança. O MAPP-CIES foi, assim, elaborado numa perspetiva compreensiva, seguindo ou promovendo uma estratégia metodológica multimétodo.

A expectativa é que o modelo proposto possa ser utilizado como referência nas avaliações dos programas cofinanciados pelos fundos europeus, sendo munido de versatilidade e flexibilidade, para se ajustar e dar resposta a diferentes objetos e contextos de avaliação. Nesse sentido, a componente empírica foi baseada em dois estudos de avaliação previamente realizados, nomeadamente o Estudo de Avaliação do Contributo do PT2020 para a Promoção do Sucesso Educativo, Redução do Abandono Escolar Precoce e Empregabilidade dos jovens para o POCH (POCH), e o Estudo de Avaliação de Impacto dos Contratos Locais de Desenvolvimento Social (CLDS). Estes dois estudos permitiram, primeiramente, realizar um exercício detalhado de revisão crítica dessas avaliações, fundamental para o desenvolvimento teórico do modelo e, posteriormente, serviram de base empírica para a testagem do modelo de avaliação proposto.

O plano de trabalhos seguido, visou o desenvolvido do MAPP-CIES que pretende ser um contributo para futuros avaliadores de políticas públicas complexas que inclui e propõe, em cada dimensão de análise, as questões a responder, a estratégia metodológica a seguir e os modos de contornar alguns dos constrangimentos que geralmente limitam, por um lado, o acompanhamento dos programas e, por outro lado, os estudos de impacto.

Descrição do problema

Construir, ou reconstruir, uma Teoria da Mudança (TdM) é sempre um grande desafio, mesmo para os avaliadores mais experientes. Apesar da TdM não constituir um instrumento ou abordagem recente, estando relativamente institucionalizada na prática avaliativa, parece ainda atravessar por um processo de consolidação e de afirmação. Ela é comumente mobilizada em avaliações ex-ante e ex-post, mas também utilizada para a programação de políticas, projetos ou programas. Por um lado, é tida/apresentada como um instrumento/ferramenta metodológica e, nessa medida, concorre com outras abordagens que testam resultados numa avaliação (como a Análise de Impacto Contrafactual); por outro lado, é incorporada no exercício avaliativo com uma latitude mais abrangente, constituindo uma abordagem com estatuto teórico-metodológico que visa não apenas a testagem dos resultados finais da avaliação, como ainda orienta teoricamente todo o processo de avaliação, tornando relevantes resultados intermédios e outras fases de avaliação e da vida de uma política pública. É nossa convicção que o modelo MAPP-Cies recai nesta última abordagem.

O desenvolvimento teórico-conceitual da TdM no âmbito do MAPP-CIES é o resultado do cruzamento da revisão de literatura e a análise documental com o trabalho empírico realizado, nomeadamente a análise dos estudos de casos e das entrevistas, que orientaram o enfoque analítico e suportaram as opções de investigação.

As entrevistas possibilitaram, entre outros aspetos, identificar e compreender as principais dificuldades e desafios associados ao processo de elaboração de uma TdM e realização do exercício de Análise de Impacto Contrafactual (AIC). Tal enquadramento terá permitido a calibragem do MAPP-CIES, sobretudo a reflexão sobre a sua capacidade para providenciar respostas e soluções.

Assim, interessa aqui dar a conhecer as principais conclusões retiradas da análise empírica, que sumariamos em alguns pontos:

- I. Depreende-se a partir das auscultações realizadas que a TdM é um assunto ainda fortemente debatido nos dias de hoje, não existindo uma ideia consensualizada sobre o papel que esta (deve) desempenha(r) numa avaliação de uma política pública.
- II. Não existem fórmulas ou métodos únicos que visem a sua construção e ou interpretação. É certo que esta flexibilização podia trazer vantagens de adaptabilidade, conferindo maior elasticidade ao exercício de (re)construção de uma TdM, mas acaba por acrescentar dúvidas no avaliador e por determinar o aumento da exigência e complexidade dos cadernos de encargos, cada vez mais inflexíveis e com orientações mais formatadas sobre o modo como a TdM deve ser perspectivada e ou apresentada.
- III. Mais recentemente, os cadernos de encargos parecem também balizar quase toda a metodologia a usar na avaliação (exigem técnicas, formas de apresentação da informação, formas de tratamento). Aparentemente essa parametrização parece facilitar o trabalho do avaliador, mas na verdade pode espalhar a elaboração de uma TdM.
- IV. Do ponto de vista prático, a falta de elementos que explicitem a TdM desde o momento inicial do programa, ou a dificuldade de acesso aos programadores da política, dificulta o processo de reconstrução da TdM numa avaliação ex-post. É particularmente difícil recuperar o racional que está por trás do

programa e a sua lógica de funcionamento. Ou seja, tal significa que o avaliador pode não conseguir chegar às reais preocupações que justificaram a intervenção.

- V. Com alguma frequência existem queixas sobre a forma como a TdM se articula com as questões de avaliação. Por vezes, existe a tendência para concentrar os esforços de interpretação de uma TdM apenas nos domínios da eficácia ou do impacto. O MAPP-CIES mostra como não cair nessa centralidade, tornando também relevantes resultados intermédios e outras dimensões de avaliação (Ver guião Anexo X).
- VI. Verifica-se algum desencontro entre os recursos e a informação a recolher necessária para o desenvolvimento de um exercício robusto e exaustivo de construção da TdM e o tempo/orçamento disponibilizado às equipas de avaliação. A elaboração de uma TdM envolve, indesejavelmente, um curto espaço de tempo, com pouco tempo reservado para reflexão. Sob este ponto de vista, e de forma global, o exercício de (re)construção da TdM torna-se ineficiente e pouco produtivo.
- VII. Por vezes o objeto de avaliação é demasiado amplo, sendo difícil o encontro de um racional que aproxime as várias medidas que integram o programa em avaliação. Isso traz dificuldades acrescidas no âmbito da elaboração de uma TdM, complexificando, sobretudo, a fase de testagem e de interpretação do encadeamento dos mecanismos.
- VIII. Por vezes as conclusões têm uma base de sustentação mais incipiente. Ou seja, são sustentadas parcialmente por alguns dos resultados, ou pela visão de apenas alguns dos stakeholders auscultados, ou estão ainda muito marcadas pela própria experiência e mindset do avaliador. Isto pode significar que a TdM foi em algum momento contornada, dificultando também a sustentação das recomendações usualmente redigidas no final da avaliação. Além das questões/dimensões de avaliação mais comuns, recorrer à elaboração de várias subquestões que orientem a avaliação a partir do encadeamento da TdM, tal como proposto pelo MAPP-Cies, pode ajudar a não cair nestas armadilhas.
- IX. A AIC é entendida como um método complementar à TdM, e por si é insuficiente para a avaliação de políticas públicas. Esta complementaridade é entendida face à característica quantitativa inerente ao método estatístico.
- X. Do ponto anterior, foram assumidas que subsistem algumas divergências de entendimento quanto ao estatuto da AIC na avaliação de políticas públicas.
- XI. Foram reconhecidas limitações à realização da AIC, nomeadamente a sua não aplicabilidade a todas as tipologias de intervenções, a inexistência de dados requeridos, a dificuldade em aceder a certos dados por motivos dos regulamentos de proteção dos mesmos, abrangência e validade dos resultados do próprio exercício.
- XII. Face aos constrangimentos identificados, admite-se a necessidade de construir um diálogo entre avaliadores e entidade contratante, que resulte num entendimento de execução ou não da AIC, e das alternativas a mobilizar caso a AIC não seja exequível.
- XIII. As expectativas de enquadramento da AIC face ao MAPP-CIES; prendem-se essencialmente:
 - i. pela compreensão nos cadernos de encargos da necessidade de alguma flexibilidade na decisão sobre a sua execução;
 - ii. por uma orientação do estatuto da AIC no processo de avaliação;
 - iii. pela reflexão e criação, antecipadas, das condições à sua realização;
 - iv. e pela potencialização do próprio método de contrafactual como instrumento de avaliação, isto é, criar as condições de aplicabilidade a um conjunto mais alargado de intervenções e uma maior robustez interna do exercício.

Face ao exposto, o enquadramento focará a clarificação teórico-conceitual da TdM. a delimitação do seu âmbito analítico, a sua articulação com as fases de concretização e avaliação das políticas e a sua operacionalização em cada uma delas.

Referenciais Teóricos

Teoria da Mudança: definição dos principais conceitos

Como ponto de partida para a apresentação dos principais conceitos que compõem a TdM e como vai ser esquematizada no âmbito do MAPP-CIES, partilhamos uma definição de TdM proposta por Helen Cahill:

“A teoria da mudança é um esboço do processo de mudança esperado. Este identifica os objetivos e as metas, inclui a descrição do contexto e o ponto de partida. São nomeadas as estratégias que serão usadas para concretizar os resultados desejados. É providenciada a lógica explicativa a partir da qual se prevê que as estratégias selecionadas sejam capazes de causar a mudança. Integra a teoria, ou a história explicativa através da qual se espera que as estratégias propostas gerem a mudança. Esses mecanismos de mudança podem ser lineares, circulares ou rizomáticos. Podem ser singulares ou múltiplos. Usualmente incluem uma narrativa sobre as maneiras como o processo pedagógico incorporado, experiencial, exploratório, performativo e co construído contribuem para a mudança. Por vezes denominado de 'modelo lógico', a teoria da mudança torna-se o documento que descreve as entradas, as saídas e os resultados.” (Cahill, 2018, p. 180)

Desmontando a definição supracitada, a TdM consiste nas teorias que estão na base do desenho e da concretização da intervenção a realizar, assumindo como primeira premissa que as políticas públicas têm uma lógica constitutiva e operacional que lhes dá forma e que orienta a sua execução, sendo que essa lógica pode estar explícita ou implícita, aprofundada ou superficial, estruturada ou fragmentada. Independentemente do seu desenvolvimento ou explicitação, as políticas são compostas por um conjunto de assunções e hipóteses de partida sobre para que servem, como podem concretizar a mudança social e que efeitos e impactos podem gerar. A segunda premissa é que estas assunções e hipóteses estruturam a totalidade da intervenção, isto é, orientam as práticas dos atores nas fases de conceção, planeamento, implementação e execução. Mais ainda, estas teorias estão articuladas na forma de encadeamento e causal e organizadas em minietapas, independentemente do seu grau de formulação e explicitação por parte do programador ou gestor, refletindo as suas crenças sobre o modo de obtenção da mudança desejada. Assim, a (re)construção da TdM como ponto de partida para a avaliação, visa providenciar um mapa da lógica processual da intervenção, baseada no quadro teórico-conceitual a partir do qual foi desenhada e implementada a política. Ou seja, visa explicitar e caracterizar as assunções e hipóteses que integram cada uma das minietapas, a forma como estas sequências de minietapas se articulam cumulativamente e se materializam em atividades e ações, que contribuem para a obtenção dos resultados desejados. A concretização destas premissas no exercício avaliativo traduz-se na terceira premissa, que estipula que as políticas públicas para serem avaliadas nos termos em que foram desenhadas e executadas têm que reconstruir a lógica racional que esteve na sua origem, conectando os inputs da política (as ações ou atividades realizadas) com os outputs desejados (os efeitos ou impactos gerados). O exercício de avaliação consiste, assim, no teste empírico das teorias e hipóteses que estiveram na base do desenho da intervenção e do modelo de operacionalização da mesma.

Assim, a primeira tarefa do exercício avaliativo que usa esta abordagem metodológica é a (re)construção da teoria da mudança associada ao programa, medida ou intervenção, explicitando as assunções dos atores e desenhando hipóteses testáveis, que vão ser validadas ou invalidadas nos processos de implementação e de execução das atividades. Para um teste eficaz dessas teorias é necessária a descrição das mesmas com o máximo de detalhe possível, identificando todas as assunções incorporadas na intervenção. Com base nesse exercício de levantamento e esquematização, os avaliadores selecionam e/ou constroem métodos de recolha e análise de informação para avaliar o desdobramento das suposições e os resultados decorrentes da intervenção. O objetivo é examinar se as

teorias que estão na base da intervenção se sustentam e em que medida produzem a mudança desejada (Weiss, 1995). Tendo como referência os efeitos desejados e exetáveis da intervenção, que compõem os resultados diretos, a avaliação tem ainda que considerar os restantes efeitos e impactos gerados indiretos, que podem incidir sobre os beneficiários, os stakeholders envolvidos, a comunidade ou a sociedade como um todo, e as consequências não intencionadas (Jaffe, Crooks, & Wolfe, 2003; Oliver, Lorenc, & Tinkler, 2020), que dizem respeito aos efeitos e impactos que os programadores e restantes intervenientes não pretendiam gerar com a intervenção, podendo estes ser diretos ou indiretos.

No entanto, o exercício de avaliação não se pode limitar a reconstruir o racional lógico com que a medida foi desenhada e operacionalizada. Apesar de a abordagem da TdM estar ancorada nas assunções teóricas que constroem e compõem as intervenções, tal opção seria assumir que esse racional engloba a totalidade dos processos de mudança gerados pela concretização dessa intervenção. O desenho de uma política é construído a partir de um conjunto de assunções feitas pelo programador sobre os problemas que procuram dar resposta, sobre o diagnóstico, sobre as ações a realizar e sobre os efeitos e impactos esperados, sendo que estas assunções podem ser suportadas no conhecimento científico, no conhecimento do terreno, em experiências pessoais, em experiências externas e internacionais, etc.

Por outro lado, e independentemente do grau de consolidação do racional lógico da política e do grau de compreensão que os programadores e stakeholders envolvidos tenham sobre a mudança que estão a promover, a sua execução coloca em funcionamento um conjunto de mecanismos de mudança que vão gerar efeitos diretos e indiretos nos stakeholders e nos atores envolvidos nos processos de mudança. Por isso mesmo, é essencial que o exercício avaliativo procure explicar como é que as medidas e ações implementados devem acionar mecanismos de mudança e quais os resultados esperados da sua concretização. O conhecimento científico, ancorado nas teorias das Ciências Sociais, suportado numa abordagem multidisciplinar que convoca os conhecimentos específicos das diferentes áreas científicas pertinentes para avaliar o objeto de avaliação, vai constituir a base a partir do qual se faz esse exercício. Esta explicitação vai permitir, por um lado, obter uma visão complexa e consolidada da realidade e da mudança que se pretende gerar, e, por outro lado, construir um referencial para a avaliação dos resultados obtidos, que pode ou não coincidir com os resultados desejados.

Após a reconstrução da lógica subjacente à intervenção, as assunções e hipóteses dos atores têm que ser validadas através do recurso ao conhecimento científico sobre as problemáticas em questão. Se a lógica causal permite-nos conhecer as assunções e hipóteses dos intervenientes e compreender as suas práticas na sua participação na intervenção, o conhecimento técnico e científico permite-nos aprofundar e complexificar o que sabemos sobre a realidade que se pretende mudar, sobre os fatores-chave e os mecanismos a ativar e quais os resultados e os impactos esperados, designadamente os benefícios pretendidos, os custos associados, as oportunidades geradas e os riscos potenciais. As ciências sociais permitem também enquadrar e compreender as assunções dos atores no desenho e execução da intervenção, identificando potenciais desfasamentos face ao conhecimento científico ou técnico, pré-conceitos, ideologias explícitas ou implícitas, que serão importantes para a análise e avaliação das práticas e dos resultados.

Por outro lado, a lógica causal que está na base do desenho e da construção da intervenção deve ela própria constituir matéria de avaliação. Um processo avaliativo que se baseie apenas nas interpretações dos intervenientes aceita implicitamente o desenho da intervenção e os processos que estiveram na sua origem. Para a produção de melhores políticas públicas, a fase de conceção das intervenções deve ser avaliada nos seus processos e produtos. Ou seja, não importa apenas aferir se o desenho da medida e as teorias implícitas são adequadas, mas também se os processos que estiveram na sua base foram fundamentados, consistentes e rigorosos. Neste âmbito, as ciências sociais providenciam instrumentos que permitem avaliar o diagnóstico realizado, as necessidades identificadas, as estratégias desenhadas, a articulação e coerência entre as atividades, ou os resultados esperados.

Em muitas situações as TdM que orientam os programas não são explicitadas pelos programadores ou operadores. Isso não quer dizer que não existam, ou que não sejam partilhadas, mas sim que existem num plano implícito. Nessas situações, o avaliador precisa de reconstruir a lógica da intervenção nas suas várias fases de concretização, desde a identificação do problema aos resultados finais, por forma a orientar o processo de recolha e análise de informação empírica e a estabelecer um referencial que permita contextualizar os resultados recolhidos. A necessidade de construção da TdM não torna inexequível o trabalho avaliativo, mas torna o exercício mais complexo e demorado. Exige um processo de reflexividade dos atores e stakeholders envolvidos sobre o seu trabalho e como influenciaram a sua participação e os efeitos que procuraram promover, bem como a capacidade dos avaliadores de articularem as teorias fragmentadas num modelo lógico integrado.

A construção da TdM é também indispensável quando esta se encontra explicitada, seja por quem concebeu e desenhou a medida ou por uma avaliação ex-ante. O avaliador não pode assumir que a explicitação da TdM é conhecida e partilhada pelos vários intervenientes e stakeholders ao longo das várias fases de concretização da intervenção. A concretização da teoria da mudança não depende apenas da visão de quem concebe e desenha a medida, mas também dos gestores das intervenções, dos stakeholders institucionais, dos técnicos no terreno e dos destinatários-alvo. O processo de mudança gerado pela intervenção composta da multiplicidade destas perspetivas sobre o problema, a intervenção e os resultados ambicionados.

Para adensar a complexidade do exercício, as assunções que orientam o desenho, planeamento e execução do programa podem ser múltiplas, na medida em que os diferentes participantes podem ter teorias tácitas distintas sobre o que é e como concretizar o programa, direcionando a sua ação e atuando de forma divergente ou eventualmente conflituosa com a visão de quem desenhou o programa (Weiss, 1995). Assim, uma intervenção, bem como a sua avaliação, pode operar com múltiplas teorias, antecipando diferentes processos a partir dos quais podem ser atingidos os resultados desejados. A implicação para o exercício avaliativo é a maior pertinência de levantamento das teorias que orientam a lógica causal do programa, explicitando as assunções, definindo os métodos de operacionalização, clarificando os mecanismos de mudança e estabelecendo os objetivos a atingir. Avaliar a concretização minietapas ao longo de toda a intervenção, seguindo todo o ciclo de vida do programa, torna mais plausível que os resultados sejam devidos às atividades do programa e não a eventos ou fatores externos da avaliação.

Um estudo raramente pode recolher dados sobre todas as assunções e hipóteses que estão presentes na construção da intervenção. O exercício de acompanhamento e avaliação de todas estas teorias tornaria a avaliação excessivamente complexa e inoperacional. Nesse sentido, o avaliador tem que fazer escolhas sobre quais das várias teorias devem ser mobilizadas para o exercício avaliativo. Para a seleção das teorias mais pertinentes destacam-se três critérios principais. O primeiro critério é o grau de influência das teorias. Por exemplo, teorias sobre a intervenção enunciadas pelos programadores ou gestores das medidas ou que são partilhadas pela globalidade dos atores envolvidos têm um elevado grau de influência. O segundo critério é a centralidade das teorias, ou seja, em que medida é que as assunções e teorias dizem respeito ou têm implicações decisivas para os resultados obtidos com a intervenção, sendo que a definição dessa centralidade é uma decisão do avaliador, baseado na sua interpretação dos aspetos centrais do modelo lógico de intervenção e apoiado no conhecimento científico sobre a mesma. Por exemplo, mesmo uma teoria pouco partilhada pode revelar-se pertinente seguir dada a sua centralidade para o sucesso ou insucesso da intervenção. O terceiro critério é a plausibilidade das teorias. Na análise preliminar das teorias recolhidas sobre o processo de mudança, o avaliador tem que aferir se estas são passíveis de responderem aos objetivos da intervenção. O conhecimento científico constitui um recurso valioso para apoiar o processo de aferição da plausibilidade das teorias (Weiss, 2000).

Para uma análise e seleção adequada das teorias mais adequadas ao exercício avaliativo em questão, afigura-se como fundamental o cruzamento de fontes de informação, nomeadamente a revisão de literatura e a realização

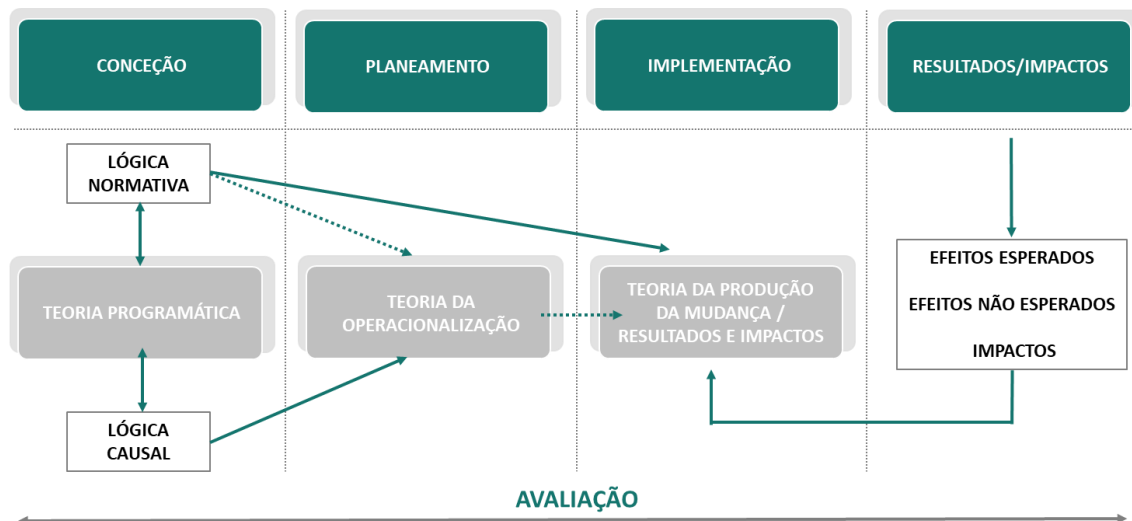
de entrevistas com especialistas, bem como a articulação das interpretações científicas recolhidas com as teorias causais e a realidade concreta intervencionada. De notar que o conhecimento científico também pode conter múltiplas abordagens e perspetivas contraditórias, não devendo ser abordado pelo avaliador como uma receita técnica consensualizada para a intervenção na realidade intervencionada. Pelo contrário, deve incorporar este conhecimento em toda a sua complexidade e dialética, confrontando paradigmas teóricos e resultados diferenciados.

Em suma, a construção ou a validação da teoria da mudança parte de um conjunto de questões de partida que se coloca ao programa ou à intervenção a serem avaliados, a saber: O que é que os programadores esperam realizar? Qual o contexto, os problemas e os objetivos que se pretendem atingir? O que deve ser feito em teoria para concretizar essa mudança? Quais são os recursos necessários para a implementação das atividades e ações que conduzem à mudança? Quais os arranjos institucionais e normativos que potenciam a concretização dos objetivos pretendidos? Quais são os sinais prováveis de que essas mudanças aconteceram? A lógica de conceção e implementação do programa ou intervenção é então desenvolvida de trás para frente a partir dessa visão de mudança que vai orientar o processo avaliativo, por um lado, estabelecendo o quadro interpretativo a partir do qual a medida ou programa vão ser avaliados, e, por outro lado, definindo os termos, ou resultados esperados, que vão permitir aferir de que forma a mudança pretendida foi ou não concretizada.

Operacionalização da Teoria da Mudança

No âmbito do MAPP-CIES, propomos operacionalizar a TdM em todas as fases que perfazem a vida de uma política pública. Na clarificação conceptual foi argumentado que as políticas públicas são construídas a partir de assunções e hipóteses sobre o processo de mudança pretendido, são organizadas e articuladas de forma cumulativa em minietapas, que vão estabelecer o caminho e a lógica do programa ou intervenção.

Figura 1: Esquema geral da Teoria da Mudança



Nesse sentido, a TdM não diz respeito apenas à fase de conceção das medidas de política, mas a todo o ciclo de vida da intervenção, incluindo assim as fases de planeamento, implementação e os resultados e impactos previstos. Por isso mesmo, a TdM não se refere apenas aos mecanismos de mudança que incidem diretamente sobre os destinatários-alvo das intervenções, mas também a todo o processo de planeamento e implementação das intervenções. Por outro lado, as avaliações nem sempre incidem sobre a totalidade dos programas ou intervenções. O cliente pode pretender avaliar fases específicas do programa ou intervenção, o que não invalida a aplicação desta

abordagem. Pelo contrário, torna-a ainda mais pertinente porque permite enquadrar e contextualizar cada fase com a totalidade da intervenção.

Respondendo aos objetivos de desenvolvimento de um modelo de TdM capaz de providenciar uma grelha teórico-metodológica (Figura 1) que permita analisar simultaneamente a totalidade da intervenção e as especificidades de cada uma das suas fases, propomos a sua operacionalização em três níveis de elaboração, a saber:

1. Teoria Programática: incide sobre o desenho global do programa e como organiza as medidas em função dos problemas e das necessidades, quais os objetivos que são esperados e que medidas, ações ou iniciativas vão ser concretizadas para atingir a mudança.
2. Teoria da Operacionalização: incide sobre a operacionalização do programa ou medida, ou seja, sobre a forma como vai ser concretizado no terreno. Visa extrair qual a teoria subjacente ao planeamento e modelo de implementação definido para o programa: quais os stakeholders institucionais e outros atores envolvidos, qual o estatuto atribuído a cada um, o modelo de distribuição de recursos, as atividades criadas, a definição dos destinatários-alvo, a estratégia de comunicação, o modelo de monitorização, entre outros.
3. Teoria da Produção de Efeitos: incide sobre a execução das ações ou a aplicação das medidas, ou seja, sobre a forma como a sua concretização no terreno vai despoletar e efetivar processos de mudança, colocando em relação os atores e stakeholders que aplicam as medidas e os beneficiários finais a quem estas são dirigidas. Os mecanismos de produção de efeitos são entendidos como a combinação das atividades realizadas, com os pressupostos que lhe estão associados, com as respostas geradas por estas nos beneficiários das mesmas.

Socorrendo-nos da proposta de Chen (1990) e de Chen e Rossi (1983), a teoria programática remete para os fundamentos e medidas que estão na base da política ou do programa. A construção da teoria programática visa explicitar a teoria que integra o problema, articulando numa sequência faseada de causas e efeitos, as necessidades, os objetivos, a estratégia de intervenção, os recursos mobilizados, as ações a realizar e os resultados esperados (Rogers, 2008; Rogers, Petrosino, Huebner, & Hacsí, 2000). Por outras palavras, com a teoria programática pretendemos saber o que e porque se pretende iniciar um processo de mudança, como se pretende gerar essa mudança e que efeitos e impactos se esperam gerar, ou seja, pretende-se explicitar o “porquê”, o “como” e o “para quê” da intervenção.

Na generalidade dos casos, não são diretamente visíveis ou conhecíveis pelos avaliadores em toda a sua complexidade, pelo que a sua explicitação é concretizada em dois planos que se articulam na implementação e na avaliação da medida ou programa, a saber: (i) o raciocínio causal e (ii) o raciocínio normativo. O raciocínio causal corresponde às teorias sobre como o programa funciona por parte dos stakeholders envolvidos, entre os quais se destaca quem concebeu e desenhou o programa. O raciocínio normativo corresponde à teoria suportada pelo conhecimento técnico e científico que fornece orientação sobre o funcionamento dos mecanismos de mudança e quais objetivos e resultados devem ser aprofundados ou testados.

A reconstrução do raciocínio causal visa explicitar o modelo lógico da intervenção, ou seja, quais as várias minietapas que compõem a intervenção, qual o conteúdo e objetivos de cada uma, como se articulam entre elas, que resultados são esperados, e como contribuem para a produção da mudança desejada. A reconstrução do raciocínio causal não pretende testar o conhecimento dos intervenientes sobre o fenómeno intervencionado, visa sim explicitar a sequência causal de como os intervenientes pretendem produzir a mudança. Dá conta da forma como os intervenientes envolvidos nas várias fases do programa ou intervenção interpretam o que estava a ser realizado, portanto as suposições e hipóteses pelas quais as pessoas definem o plano e as suas ações (Weiss, 1997). A reconstrução da lógica causal da intervenção permite ainda estabelecer um referencial para avaliar se a concretização prática da intervenção correspondeu do que foi projetado inicialmente pelos programadores.

Em tese, o raciocínio causal que está na origem da intervenção não tem que estar correto ou ser globalmente aceite pelo conhecimento científico ou técnico para ser testado. Ainda assim, se a avaliação determinar que os pressupostos teóricos que estiveram na base do desenho do programa são pouco substanciados, para além dessa evidência ser um resultado da avaliação em si mesmo, os avaliadores devem considerar teorias alternativas sobre o processo de mudança na testagem prática da intervenção, sendo que para esse efeito o recurso à reconstrução do raciocínio normativo é particularmente útil.

O raciocínio normativo aborda a pertinência e construção do programa com base no conhecimento científico e técnico, por forma a poder explicar porque e como as ações desenvolvidas podem gerar processos de mudança e quais os resultados que podem ser esperados na sua concretização, estabelecendo assim as referências a partir das quais os mecanismos de mudança são avaliados. A construção da lógica normativa da intervenção permite avaliar se este modelo lógico de intervenção é ou não capaz de gerar a mudança desejada e antecipar e explicar os efeitos e os impactos gerados pelo processo de mudança.

A incorporação do raciocínio normativo no exercício avaliativo decorre do reconhecimento de que as intervenções têm características intrínsecas e a sua concretização tem efeitos concretos nos destinatários diretos e indiretos, que vão para além das intenções e expectativas dos intervenientes. As interpretações dos intervenientes influenciam positiva ou negativamente a potenciação das mesmas, mas não esgotam os efeitos gerados. Por exemplo, numa intervenção baseada em formação profissional, independentemente da interpretação dos formadores do seu papel e dos resultados da atividade, a existência da formação estabelece um contexto de transmissão de conhecimento e de aprendizagem que vai gerar efeitos nos destinatários. Portanto, se interessa ao avaliador perceber como é que os intervenientes interpretaram a intervenção, por forma a entender o que foi feito, porque foi feito e com que expectativas, compreendendo os resultados obtidos, interessa também recorrer ao conhecimento científico, técnico, ou a experiências anteriores ou internacionais, para entender os princípios das intervenções baseadas na formação profissional, que mecanismos geram, o que esperar dos seus resultados, por forma a avaliar as teorias dos intervenientes e contextualizar os resultados.

No âmbito da avaliação é absolutamente necessário construir uma teoria programática que integre estas duas valências. Na medida em que cada uma das modalidades de raciocínio têm funções e âmbitos diferentes, ainda que comuniquem entre si, o avaliador deve ter cuidado em não as confundir no processo de explicitação da teoria programática. O exercício de elaboração do raciocínio normativo não pode ser a mera expressão das teorias e das hipóteses dos programadores das medidas ou outros intervenientes, mesmo que estes sejam especialistas técnicos ou académicos. Por outro lado, para entendermos as práticas dos atores não lhes podemos atribuir uma intencionalidade externa, gerada por teorias e hipóteses recolhidas ao conhecimento científico que não sabemos terem sido mobilizadas ou incorporadas pelos intervenientes. Cabe ao avaliador explicitar a lógica causal da intervenção assente na perspetiva dos autores e dos protagonistas da intervenção, e uma lógica normativa que explicita quais os efeitos e os resultados esperados da intervenção a partir do conhecimento técnico e científico.

A teoria da implementação visa reconstruir o racional estratégico que esteve na base da elaboração do modelo de conceção da política. Parte do pressuposto teórico de que se o programa for conduzido conforme o planeado, com suficiente fidelidade, qualidade e envolvimento dos stakeholders, estarão estabelecidas as condições ideais para que os processos de mudança produzam os resultados desejados. Em termos práticos, descreve as operações que têm que ser realizadas, os recursos mobilizados e as condições organizacionais, explicitando as assunções e crenças que explicam as opções relativamente à forma e aos recursos necessários para a implementação bem-sucedida da intervenção (Weiss, 1997). Este nível de elaboração da TdM assume como premissa que o planeamento das medidas ou programas é feito com base em teorias ou assunções sobre o processo de implementação, os stakeholders envolvidos, os recursos necessários a mobilizar, os mecanismos de distribuição de papéis e recursos, ou os destinatários a envolver, entre outras. Visa também reconhecer as dinâmicas institucionais, designadamente

as suas estratégias, contributos e ações dos stakeholders. As intervenções não são meras expressões técnicas de ideias do programador, são também o resultado de processos do envolvimento e da negociação destas propostas com os stakeholders de vários níveis, nos quais as suposições e hipóteses destes sobre os processos de mudança são (mais ou menos) incorporados nas propostas finais. Outra premissa na base da teoria de implementação é que a fase de planeamento não deve ser vista como neutra ou reduzida a uma lógica tecnocrática. Pelo contrário, é nesta fase que as propostas e teorias que constituem as políticas são operacionalizadas em atividades e ações concretas no terreno, estabelecendo assim as condições para a execução dos processos de mudança. O sucesso de um processo de mudança não está assente apenas na qualidade das teorias orientadoras, mas também da capacidade de concretização dessas teorias no terreno de uma forma que lhe permita atingir os objetivos previstos e a mudança projetada.

A teoria da produção de efeitos remete para a identificação e caracterização dos processos de mudança associados a cada ação ou atividade, e estabelecer as expectativas do seu potencial efeito junto dos destinatários, diretos e indiretos, estabelecendo os termos a partir dos quais os resultados serão avaliados. A explicitação da teoria da produção de efeitos visa, assim, elencar os mecanismos de mudança que integram cada ação da intervenção e são ativados na relação entre a sua execução e os efeitos gerados nos destinatários-alvo. Em termos simples, as teorias e hipóteses que explicam como é que aquela atividade vai produzir determinados efeitos nos destinatários e eventualmente gerar mudança. É nesta altura que se confronta o modelo idealizado teoricamente na teoria programática com as evidências empíricas nos destinatários da intervenção, e o que estava planeado na teoria da implementação, com aquilo que foi realmente feito e os resultados daí advindos. A teoria da produção de efeitos incide sobre as atividades e ações, mas é informada pelas teorias programática e da implementação (Davidson, 2000).

O pressuposto teórico que procura testar é que as atividades a realizar com os destinatários-alvo do programa ou medida são capazes de gerar os processos de mudança antecipados, atingindo os resultados necessários para o cumprimento dos objetivos estabelecidos. Assim, no âmbito de uma dada intervenção, e para cada uma das realizações ou atividades, o avaliador tem que explicitar os mecanismos de mudança e as suposições e hipóteses associados a cada um deles. A identificação destes mecanismos e a lógica que está na base da sua ativação permitirá uma análise compreensiva dos resultados obtidos, clarificando como é que cada uma das atividades contribui para a mudança desejada.

Para uma avaliação efetiva dos mecanismos de mudança que são ativados pelas atividades e ações das intervenções é necessário considerar outros fatores que podem influenciar os resultados obtidos. Com esta afirmação pretende-se evidenciar que os efeitos das atividades e ações realizadas podem ser realizados por outras vias que não a ativação dos mecanismos de mudança elencados na construção da teoria de produção de efeitos (Weiss, 1997). Nesse sentido, há dois tipos de fatores que têm que ser explicitados e integrados na análise dos processos de mudança, a saber: (i) os fatores contextuais; e (ii) os fatores de risco ou ameaças.

Os fatores contextuais dizem respeito às estruturas social, económicas e políticas, aos arranjos institucionais, ou ao contexto, histórico ou cultural, portanto, a enquadramentos ou dinâmicas sociais que podem contribuir ou limitar o sucesso de determinada intervenção para além do mérito específico da mesma (Pawson & Tilley, 1997). Por exemplo, num contexto de recessão económica profunda, como aconteceu durante a crise financeira, o sucesso das medidas de emprego é limitado, mesmo que a avaliação destas medidas ao nível da sua pertinência, adequação estratégica e implementação no terreno seja muito positiva. As intervenções estão sempre intrinsecamente relacionadas com o seu contexto de produção e implementação. Decorre daqui que as medidas e programas não são diretamente transferíveis em função dos resultados obtidos. Para garantir a transferibilidade é necessário avaliar quais e com os fatores contextuais contribuíram para as medidas e averiguar se estão reunidas as mesmas condições no novo contexto de implementação.

Os fatores de risco ou ameaças decorrem das condições necessárias para que os mecanismos de mudança sejam eficazes. Há um conjunto de pressupostos que têm que ser cumpridos para que as estratégias possam ser aplicadas tal como foram projetadas, sendo que alguns destes pressupostos possam incluir alguns riscos intrínsecos ou extrínsecos, que podem afetar as condições de implementação. No desenho e planeamento das intervenções devem ser elencados quais são os fatores que podem colocar em risco o sucesso da intervenção. Os fatores de risco podem ser de cariz contextual ou operacional. Por exemplo, do ponto de vista contextual, as medidas de confinamento durante a pandemia Covid-19 e as necessidades de ajustamento em função da propagação do vírus e do número de infetados, representou um contexto de risco para a implementação de medidas ou programas em função da instabilidade nas condições de implementação das atividades ou ações no terreno. Outros exemplos de cariz operacional, são a capacidade de mobilizar determinados stakeholders ou os destinatários-alvo para a concretização das estratégias planeadas, ou se os técnicos no terreno partilham das mesmas assunções e hipóteses para o processo de mudança dos legisladores ou se estão devidamente capacitados para a concretização das ações. Assim, o avaliador na construção das teorias da mudança para cada fase de concretização da intervenção tem que considerar os potenciais desvios ou elementos que podem ameaçar as assunções e hipóteses projetadas na construção da teoria de produção de efeitos. A utilização desta abordagem metodológica obriga a que a análise dos resultados seja feita por relação com as teorias construídas sobre o processo de mudança, pelo que a ocorrência de fatores que alteram as condições ou as assunções sobre o processo de mudança devem ser incluídos na análise dos resultados. Em tese, podemos ter uma boa teoria da mudança que não obtém os resultados esperados, mas também ter bons resultados que decorrem de desvios da teoria da mudança tal como esta foi projetada.

Finalmente, a avaliação da teoria de produção de efeitos deve ainda considerar as consequências não intencionadas e não antecipadas da intervenção (Garfield, 2004; Oliver, Lorenc, & Tinkler, 2020). Apesar da quantidade e qualidade da informação recolhida e tratada na construção da TdM, existe a possibilidade que a intervenção gere efeitos que não tenham sido previstos quer pelos atores envolvidos no processo de mudança quer pelos avaliadores. A construção da TdM não se deve impor sobre a realidade social, apenas considerando os fatores, processos e resultados previstos no seu modelo teórico e lógico. Para além disso, as consequências não-previstas podem influenciar, positiva ou negativamente, a forma como avaliamos os resultados da medida de política. Em tese, uma determinada atividade pode gerar um mecanismo de mudança que não foi previsto pelo programador ou os técnicos no terreno, mas que traz benefícios aos destinatários-alvo e por isso justificar a sua existência. Do mesmo modo, uma determinada intervenção pode ser bem-sucedida na sua implementação e gerar a mudança pretendida, mas ter associadas consequências não antecipadas negativas que podem justificar a descontinuação da mesma. Portanto, a identificação e ponderação das consequências não-previstas é fundamental para uma avaliação do impacto global de qualquer medida de política.

Importa ainda notar que os fatores contextuais e os fatores de risco não são mutuamente exclusivos. Aliás, o exercício de diagnóstico usualmente inclui o levantamento das tendências evolutivas contextuais, no qual um dos objetivos é precisamente a identificação de fatores que podem ameaçar a implementação e o sucesso da intervenção. Ou seja, os fatores de risco podem ser de ordem contextual. Da mesma forma, os fatores de risco não têm necessariamente que afetar diretamente a intervenção em avaliação, mas sim o contexto de implementação.

Os níveis de elaboração da TdM relacionam-se com as várias fases de concretização do programa ou da intervenção, visando estabelecer o modelo lógico que orientou a sua estruturação e estabelecer os parâmetros que vão permitir testar as teorias, articulando (i) as assunções e crenças que estiveram na origem e orientaram a concretização do programa, (ii) as assunções e hipóteses que estruturaram a implementação do programa ou medidas e (iii) os processos e resultados esperados no desencadeamento dos mecanismos de mudança junto dos destinatários-alvo.

A TdM constitui um modelo dinâmico e compreensivo, sequenciado em minietapas cumulativas, pelo que cada uma destas fases de elaboração é informada e influenciada pelas restantes, não se constituindo com unidades estanques e independentes. Por exemplo, as estratégias a implementar no terreno para a concretização de determinada intervenção podem ser condicionadas pelos recursos financeiros disponíveis. Da mesma forma, o tipo de stakeholders a envolver numa medida é influenciado pelos destinatários-alvo a quem a medida é destinada. Portanto, se a elaboração de cada uma destas teorias foca cada uma das fases da intervenção, a sua construção tem que considerar as lógicas e dinâmicas de articulação entre os vários níveis de elaboração da TdM.

Finalmente, a avaliação da TdM nos seus vários níveis tem que considerar não apenas a sua adequação processual, a forma como a concretização prática da teoria espelha as suposições e hipóteses que estiveram na sua base, mas essencialmente os resultados gerados pela intervenção, se os efeitos e impactos decorrentes da concretização destas teorias produzem a mudança pretendida. A avaliação de cada um dos níveis de elaboração da TdM é o resultado da relação entre a fase em que incide com os efeitos e impactos gerados. As suposições e hipóteses que compõem cada uma das minietapas que compõem as teorias da mudança são providas de sentido por influenciarem a concretização de um processo de mudança. Cabe ao avaliador, seja no processo de recolha de informação ou na análise do material empírico, articular cada uma das minietapas com os resultados gerados pelas mesmas, sejam estes finais, intermédios ou iniciais, diretos ou indiretos, e como contribuem para a mudança gerada pelo programa ou intervenção.

Análise de Impacto Contrafactual como método quantitativo

Poder-se-á afirmar que a principal vantagem da aplicação da AIC é a possibilidade de estimar os resultados e os impactos de uma determinada política pública, permitindo responder em que medida (quanto?) o programa produziu mudança. Contudo, a análise estatística contrafactual não é condição per se para a atribuição da relação de causalidade entre a intervenção e o resultado. De facto, nem sempre é possível aplicar um método experimental, ou seja, um ensaio de controlo aleatório também conhecido como “experimentação social” (Comissão Europeia, 2013). Em alternativa, a AIC pode fundar num método quase experimental, de abordagem não aleatória, em que apenas é possível controlar alguns fatores que influenciam o resultado (este é um primeiro tipo de “irrealidade” convocada pelo método). Ora, a construção dos modelos de contrafactual partem sempre de algumas suposições teórico-metodológicas, não só associadas à técnica de contrafactual especificamente adotada, mas também à natureza substantiva do problema. Identificar e estimar os efeitos causais requer sempre algum nível de teoria, nomeadamente no momento da escolha das variáveis de causa e de efeito, em particular teoria científica (Heckman, 2005). Por outras palavras, as escolhas técnico-metodológicas na análise contrafactual têm de ter uma sustentação substantiva. Tal como numa relação estatística de associação, onde é preciso demonstrar que a correlação não é espúria, também na relação de causalidade é necessário sustentar a atribuição do efeito da intervenção a uma teoria. Adicionalmente, a leitura dos resultados e dos impactos quantificados pela AIC só tem expressão quando substantivamente interpretados e enquadrados. Além da AIC dever estar ligada a um quadro substantivo, como uma teoria científica, acredita-se que pode ser potencializada quando devidamente enquadrada num quadro mais alargado de avaliação de políticas públicas, nomeadamente num modelo de avaliação baseado na teoria, em concreto a teoria da mudança.

Análise de Impacto Contractual e Teoria da Mudança

Se, por um lado, a AIC responde à questão de saber em que medida o impacto pode ser atribuído à intervenção, por outro lado, é incapaz de responder “como” e “porque” é que a intervenção contribuiu para a mudança. Por exemplo, a partir da avaliação dos cursos profissionais do PO CH (2021), concluiu-se que a diversificação da oferta formativa para uma maior componente prática e de contacto direto com o mercado de trabalho é um dos mecanismos eficazes a contribuir para a redução da desistência escolar no secundário, o “como”, pois essa componente prática aumenta a motivação e a satisfação dos alunos nos estudos e a valorização do papel da escola

nas suas possibilidades profissionais, o “porquê”. As respostas a estas duas últimas questões, o “como” e o “porquê”, só se divisam por intermédio da mobilização de um quadro teórico, pelo que a teoria da mudança e a AIC não são exclusivos numa avaliação. Se há quem considere a avaliação baseada na teoria (ABT), onde se inclui a Teoria da Mudança, como uma alternativa à avaliação de impacto na ausência e impossibilidade de uma análise contrafactual quantitativa (Bamberger, M., Carden, F., & Rugh, J., 2009; Hind, 2010), numa outra perspetiva há quem defenda que ambas são complementares, pois respondem a questões distintas (White, 2009). Porém, responder a questões distintas não implica fronteiras rígidas entre as evidências que a TdM e a AIC expõem. Apesar de autónomas é possível ir além da sua complementaridade. A ideia é a de que a análise de impacto contrafactual é uma ferramenta, entre outras, ao serviço da avaliação realista baseada na Teoria da Mudança, a que se subordina, já que esta permite validar a pertinência do exercício, selecionar as variáveis, delimitar o alcance e validade dos resultados e, ainda, interpretá-los, num quadro que não se pode reduzir à avaliação da produção de efeitos.

Como já referido, o contrafactual é um processo imaginativo, que envolve a questão “e se?”. Além das questões já mencionadas, “Foi X que causou Y?”, “Se eu fizer X, Y acontece?”, “E se eu fizer diferente?”, também as questões “Como aconteceu?” e “Porquê?” fundam sobre um raciocínio lógico de contrafactual. Quando uma avaliação tenta responder “como” a intervenção produziu o respetivo resultado, e “porquê”, está sempre implícito um cenário alternativo que ao acontecer teria produzido outros resultados. Por outras palavras, o contrafactual é um exercício lógico do cenário alternativo, “e se?”, que engloba questões de “como”, “porquê” e “quanto”. Por esta razão, o método estatístico contrafactual, que mede o “quanto”, é mais que complementar ao método baseado na teoria. Apesar de autónomos não são independentes, o que sugere que a melhor estratégia é a de articulação.

Face ao exposto, no âmbito do MAPP-CIES é importante equacionar se a análise de impacto contrafactual estatística é um método que apenas permite avaliar os resultados e os impactos, ou seja, se apenas é útil no final da cadeia dos quatro modelos de avaliação. Além de contribuir para o modelo de avaliação de resultados e impactos, poderá a AIC contribuir para os modelos precedentes na cadeia de avaliação de uma política pública?

De facto, a avaliação dos efeitos, seja qual for o método, funciona como um teste às teorias da conceção e da programação e à qualidade da implementação, pelo que além do exercício contrafactual estatístico responder à questão do “quanto”, também pode auxiliar a responder a outras questões de avaliação. Por outras palavras, a visão aqui apresentada é a de que a TdM tem uma função orientadora da construção do exercício de contrafactual, e a de que a AIC funciona como um teste à TdM, pelo que os seus resultados contribuem para a avaliação com base na teoria, incluindo a reformulação da própria teoria em que se baseia o programa.

Teoria da Mudança como “função de comando” da Análise de Impacto Contrafactual

Considere-se primeiro o caso em que a Teoria da Mudança comanda o exercício de AIC:

1. Quando desprovido de uma teoria o exercício contrafactual não encontra enquadramento substantivo para as escolhas técnico-metodológicas. A natureza do problema em análise pode influenciar o método de contrafactual adotado, e as opções técnicas inerentes ao respetivo método. Essa escolha envolve um conhecimento teórico e empírico do fenómeno em estudo, e a correta fundamentação das opções tomadas, e pode ser baseada na reconstrução da teoria do programa efetuada pela Teoria da Mudança em fase de modelo de avaliação da conceção. Adicionalmente, a validade do exercício contrafactual estatístico deve ser discutida com base nas limitações técnico-metodológicas com que o avaliador se deparou ao longo do processo;
2. Em segundo lugar, a Teoria da Mudança enquadra os resultados obtidos através da AIC. Interpretar e enquadrar a grandeza dos resultados e impactos alcançados e os desvios face às metas desenhadas é um procedimento potencializado com recurso aos resultados de avaliação preconizados pelo modelo de operacionalização;

3. A Teoria da Mudança também enquadra a interpretação dos resultados da AIC na possível identificação de mecanismos funcionais ou não, corretos e incorretos, face às descobertas alcançadas durante a fase de avaliação do modelo de produção da mudança;

4. O contrafactual estatístico é um exercício forte em termos de validade interna, mas fraco em termos de validade externa. Quer isto dizer que os resultados são válidos para o universo ou amostra para a qual se efetua o exercício, mas dificilmente generalizáveis. Nada indica se os resultados da intervenção serão igualmente alcançados num outro universo ou contexto. Uma possível extrapolação dos resultados para outros contextos ou universos só é verosímil através de uma análise qualitativa de compreensão e avaliação substantiva dos mecanismos. Deste modo, a interpretação dos resultados da análise contrafactual estatística é potencializada através da avaliação qualitativa.

É, portanto, relevante fundar parte da AIC nos desenvolvimentos da avaliação preconizados pela Teoria da Mudança.

Análise de Impacto Contrafactual como teste da Teoria da Mudança

Considere-se o segundo caso, em que os resultados da AIC são integrados no processo de avaliação com base na Teoria da Mudança:

5. Quando se constrói um modelo contrafactual o que realmente está em causa é a eficácia dos mecanismos de mudança (Pearl, 2018). Na ausência de impacto a interpretação é a de que não subsistem mecanismos a intermediar a relação entre A e B, ou que algo bloqueia o mecanismo. Logo, para o específico resultado em análise na AIC, um impacto nulo implica que a teoria causal, e a cadeia de efeitos não funcionou como expectável, ou que algo a bloqueou, o que é útil para a validação da avaliação preconizada no modelo da produção da mudança. Tal não coloca em causa a teoria do programa no seu todo, mas sim a teoria e a cadeia de efeitos específicas associadas ao resultado em medição. Note-se que apesar de análise contrafactual poder validar ou não a existência de mecanismos, nada nos diz sobre os mesmos, ou seja, o “como” e o “porquê”.

6. A análise contrafactual estatística pode ir além do impacto global médio. Dependendo da técnica aplicada, poderá ser efetuada uma análise por subgrupos de beneficiários, que mede a heterogeneidade do impacto (Comissão Europeia, 2020a; Csillag, Krekó, e Scharle, 2020; Gertler et al., 2018; Guo, e Fraser, 2015; White, 2009). Esta avaliação deve ser realizada sempre que viável, pois permite saber distinguir as ordens de grandeza dos resultados e dos impactos entre os beneficiários. Adicionalmente, também pode auxiliar a avaliação em outros três pontos: 1) na definição do grupo de beneficiários; 2) na reconsideração dos mecanismos de mudança; e 3) na avaliação do custo-benefício.

Em primeiro lugar, ao se conhecer em que subgrupos a intervenção tem maior impacto e quem mais beneficia com a intervenção, é possível avaliar até que ponto o grupo-alvo foi corretamente desenhado na fase de Planeamento, ou seja, a AIC também é útil para o modelo de operacionalização. Assim, a AIC contribui para avaliar se o grupo-alvo da intervenção e os critérios de elegibilidade na intervenção são ou não os adequados, e averiguar a assertividade dos fatores de discriminação das medidas. Por outro lado, a heterogeneidade do impacto pode simplesmente validar a teoria do programa, quando a heterogeneidade já tinha sido sinalizada durante as fases de conceção e ou planeamento, sendo por isso expectável;

Em segundo lugar, a constatação da existência de subgrupos de beneficiários onde o impacto é nulo, reduzido, ou menor pode significar que por algum motivo os mecanismos não funcionam para esses subgrupos. Poderá revelar-se importante compreender porque alguns mecanismos são mais ou menos eficazes, em particular em determinados subgrupos, o que poderá ser relevante para a avaliação elaborada no modelo da produção da mudança. Logo, é possível avaliar se se deve ou não reconsiderar os mecanismos de mudança, e se se averigua o potencial de estratégias de resposta distintas consoante diferentes destinatários. Mais uma vez, só uma avaliação

baseada na teoria, recorrendo a outros métodos e técnicas de pesquisa, permitirá responder e compreender o “como” e o “porquê”, através da avaliação da cadeia causal;

Em terceiro lugar, a análise de heterogeneidades pode ser relevante na avaliação do custo-benefício das intervenções. Subgrupos onde a intervenção não produz impacto, ou produz impacto reduzido, podem contribuir para uma maior ineficiência da medida. Por outro lado, podem existir subgrupos onde o impacto é maior, mas que são marginalmente considerados no grupo de beneficiários da intervenção. Ressalva-se que a identificação de heterogeneidades de impacto entre subgrupos não permite por si, isoladamente, justificar o reajustamento do grupo de beneficiários (nem desinvestir em subgrupos de menor impacto, nem só investir em subgrupos de maior impacto). Mas sinaliza a necessidade de aprofundar a compreensão substantiva de cada um desses níveis de impacto, através de outros métodos e técnicas preconizadas através da Teoria da Mudança. Posteriormente, caso se conclua substantivamente benéfico, poderá considerar-se a aplicação de mecanismos distintos entre subgrupos ou proceder-se ao ajustamento dos critérios de elegibilidade como forma de melhorar o custo benefício do programa (Csillag, Krekó, e Scharle, 2020);

Em suma, para compreender a grandeza dos impactos ou a sua heterogeneidade, a AIC pode contribuir para a definição de alguns métodos e técnicas de pesquisa da Teoria da Mudança, como, por exemplo, a definição de métodos e técnicas qualitativas, questões a serem incluídas em métodos qualitativos, e seleção de entrevistados. Em última instância os resultados obtidos pela AIC e pelas eventuais análises subseqüentes, caso efetuadas, podem conduzir à reformulação da teoria subjacente aos programas. Este facto evidencia a importância da realização da AIC de modo atempado no processo de avaliação. Por esta razão, é importante que a AIC inicie o seu planeamento antes da conclusão da intervenção, desejavelmente na fase de Planeamento. Durante essa fase, a previsão da realização da AIC na Avaliação Ex-Post antecipa a necessidade de recolha de dados imprescindíveis à realização do exercício contrafactual.

Estratégia metodológica

O Modelo de Avaliação de Políticas Públicas (MAPP-CIES) segue uma abordagem teórico-metodológica assente na Teoria da Mudança (TdM) e articulada com a Avaliação de Impacto Contrafactual (AIC), propondo uma perspetiva compreensiva do exercício avaliativo e uma estratégia metodológica multimétodo. Nesse sentido, o estudo é desenvolvido em dois planos que foram concretizados em articulação: (i) a revisão dos fundamentos teóricos do modelo de avaliação, baseado na revisão de literatura e em entrevistas com peritos; e (ii) o trabalho empírico, baseado na análise de dois estudos de avaliação previamente realizados, nomeadamente o Estudo de Avaliação do Contributo do PT2020 para a Promoção do Sucesso Educativo, Redução do Abandono Escolar Precoce e Empregabilidade dos jovens para o POCH (POCH), e o Estudo de Avaliação de Impacto dos Contratos Locais de Desenvolvimento Social (CLDS), e complementado com entrevistas com observadores privilegiados e stakeholders institucionais.

Assim, a abordagem metodológica desenhada para projeto MAPP-CIES visou acentuar a sua versatilidade, flexibilidade e dinâmica, por forma a ajustar-se rapidamente às necessidades da investigação, à informação recolhida e analisada na componente empírica e às conclusões de investigação. Para a realização deste objetivo já foram desenvolvidas várias ações, que explicitamos nos próximos parágrafos.

a) Revisão crítica da literatura científica, traduzindo-se na identificação de bibliografia pertinente e inovadora, na elaboração de fichas de leitura e na organização de sessões de debate e desenvolvimento teórico, de forma a ampliar o entendimento sobre a forma como a TdM e a AIC têm sido teorizadas e aplicadas, bem como as potencialidades teóricas e as possibilidades práticas da sua articulação. Estas tarefas foram realizadas ao longo de todo o projeto e afiguraram-se fundamentais para o desenvolvimento crítico e sustentado do MAPP-CIES,

nomeadamente ao nível de: clarificação dos conceitos; adequação das abordagens e dimensões analíticas aos objetivos das avaliações; sinalização e identificação de estratégias de recolha e análise da informação; desenvolvimento de modelos conceptuais para robustecer e delimitar a aplicação da TdM e da AIC; e desenvolvimento de modelos teórico-práticos para a articulação da TdM e da AIC.

b) Análise aprofundada dos dois estudos de caso, nomeadamente a Avaliação de Impacto dos Contratos Locais de Desenvolvimento Social (CLDS) e a Avaliação do Contributo do PT2020 para a Promoção do Sucesso Educativo, Redução do Abandono Escolar Precoce e Empregabilidade dos jovens para o PO CH. A análise tem por base uma grelha metodológica para a verificação e compreensão dos vários passos tomados ao longo do processo avaliativo (Anexo A). O desenvolvimento da matriz de análise dos estudos de caso foi realizado com base na revisão da literatura científica, na consulta de especialistas de avaliação de políticas públicas e através da contribuição interativa dos elementos que integram a equipa de investigação. O exercício baseia-se na análise documental dos cadernos de encargos, das propostas de candidatura, dos relatórios previstos nas avaliações, bem como nos instrumentos de recolha e de análise da informação que guiaram as opções metodológicas da avaliação, as conclusões e recomendações finais. Este exercício é orientado pela perspetiva teórica desenvolvida na revisão crítica da literatura científica.

c) Preenchimento de grelha analítica (Anexo B) para a análise de conteúdo aos documentos oficiais dos estudos de avaliação selecionados. A análise de conteúdo dos dois estudos de avaliação tinha visou a comparação entre aquilo que se preconizava no MAPP-CIES com aquilo que tinha sido feito naquelas avaliações, consideradas como que uma fonte para o modelo proposto. Esta comparação permitiu o inventário das ações realizadas ao longo das duas avaliações em análise, do ponto de vista metodológico, e facilitou uma reflexão sobre potencialidades e fragilidades das opções tomadas. A análise não se cinge apenas aos relatórios e materiais produzidos pela equipa de avaliação, mas também à documentação enquadradora dos exercícios avaliativos. Afigura-se como pertinente e necessário o conhecimento dos vários elementos que suportam a construção dos programas operacionais avaliados nos documentos oficiais produzidos e/ou reconhecidos pelas entidades gestoras, nomeadamente legislação enquadradora, o diagnóstico, os objetivos, a lógica da intervenção, dados de execução, evidências de resultados, produtos, entre outros. A concretização desta tarefa é feita em simultâneo com a revisão documental dos produtos dos estudos de avaliação.

d) A organização de um plano de recolha de informação complementar junto de stakeholders institucionais pertinentes e desenvolvimento de contactos para a sinalização dos mesmos, definição das metodologias adequadas de recolha de informação, elaboração e validação dos instrumentos de recolha de informação. O plano de recolha de informação prevê um conjunto de entrevistas e a organização de focus group que irão ocorrer de forma faseada, de modo a permitir a integração da informação, a sua revisão crítica e a preparação da fase seguinte de recolha de informação. Neste âmbito, e para os dois estudos de avaliação, foram elaboradas as matrizes gerais para os guiões de entrevista às equipas de avaliação, aos representantes das estruturas dos Programas e às equipas que acompanharam e monitorizaram a execução das avaliações (Anexo C). O seu contributo foi importante do ponto de vista da reflexão sobre as principais potencialidades e limitações dos modelos desenvolvidos nos dois estudos de caso, permitindo também sinalizar dificuldades de operacionalização do modelo metodológico e soluções encontradas para as superar, recolher propostas e recomendações para o modelo de avaliação, avaliar o impacto das avaliações, designadamente os processos de transferência do conhecimento produzido para os policy-makers e restantes stakeholders. As entrevistas contribuíram ainda para um desenho do MAPP-CIES mais robusto e ajustado às realidades concretas das avaliações de programas de políticas complexos (Anexo D). Pese embora a proposta apresentada tenha previsto a elaboração de grupos focais, os mesmos não foram constituídos. A principal razão pela qual se previu fazer grupos focais, prendia-se com o aprofundamento do conhecimento sobre os estudos de avaliação utilizados. No entanto, a elaboração das entrevistas semiestruturadas, a complementar a análise documental, tornou pouco relevante a constituição dos grupos focais. Ainda de referir que, atendendo à abertura

das entrevistas e dos guiões a novas questões e informações consideradas pertinentes e trazidas pelos participantes, houve um reforço qualitativo na compreensão de conceitos, abordagens, dificuldades e aspetos a melhorar em termos de avaliação de políticas públicas. Assim, considerou-se a informação recolhida nas entrevistas suficientemente robusta, com contributos ricos relativamente aos seus objetivos e aos objetivos inicialmente previstos para constituição dos grupos focais, retirando a necessidade de se avançar com os mesmos.

e) Organização de sessões temáticas de desenvolvimento teórico-conceitual de modo a integrar os diferentes elementos analíticos da investigação. Estas sessões mantiveram uma periodicidade mensal, ajustável às necessidades da investigação, e visaram confrontar e debater aspetos ou dimensões específicas do desenvolvimento do Modelo de Avaliação, articulando as descobertas obtidas nos diversos instrumentos analíticos desenvolvidos. Este exercício tem uma componente teórico-prática, na medida em que visou a integração e articulação do conhecimento produzido e a sua respetiva validação, refutação ou determinação de aspetos a serem desenvolvidos ou aprofundados.

f) A realização de entrevistas exploratórias não estruturadas com especialistas em metodologias de avaliação de políticas públicas. Estas entrevistas foram realizadas de acordo com questões específicas que se têm colocado no desenvolvimento teórico-conceitual, sendo, portanto, realizadas à medida das necessidades da investigação. Visaram essencialmente a consulta junto de especialistas que têm desenvolvido investigação em avaliação de políticas públicas com vista a debater conceitos, testar hipóteses de exploração teórica, sinalizar limitações metodológicas, identificar soluções operacionais e recolher bibliografia pertinente. A sua realização ajustou-se às necessidades específicas do processo de desenvolvimento teórico-metodológico e visaram complementar a revisão de literatura e as sessões temáticas. Os contributos recolhidos foram mobilizados como informação de contexto no trabalho desenvolvido.

Principais conclusões

O desenho de modelos de avaliação de políticas públicas complexas, em particular de programas financiados por Fundos Europeus Estruturais e de Investimento, constituiu um exercício desafiante.

Em primeiro lugar, atendendo à complexidade da realidade que lhes dá corpo, isto é, ao tecido social composto por diferentes atores e stakeholders, com diferentes missões, agendas e recursos. De igual modo, o tecido social está em constante evolução, fruto das dinâmicas próprias dos sistemas sociais, de influências externas, acontecimentos imprevistos, mas também por intervenções das políticas públicas. Estas têm como objetivo promover mudança no tecido social de forma intencional e estratégica, visando a qualidade de vida das populações e das comunidades, a modernização das estruturas institucionais, a justiça social, a diminuição das desigualdades sociais, entre outros fins. Dada a natureza multidimensional e dinâmica do tecido social, o desenvolvimento de políticas públicas configura-se como um exercício complexo, que integra diversos eixos de intervenção, medidas, ações e realizações que permita dar conta, de forma integrada e articulada, dos problemas sociais aos quais procuram responder.

O desenvolvimento teórico-metodológico da Teoria da Mudança (TdM) no âmbito do Modelo MAPP-CIES foi orientado por dois objetivos principais: (i) clarificar a interpretação da TdM e delimitar o seu âmbito de utilização; (ii) operacionalizar a TdM nas várias fases que compõem o ciclo das políticas.

Relativamente ao primeiro objetivo, a TdM consiste nas teorias que estão na base do desenho e da concretização da intervenção a realizar, sendo estruturada em função de quatro princípios: (i) as políticas públicas têm uma lógica constitutiva e operacional que lhes dá forma e que orienta a sua execução; (ii) as teorias estruturam a totalidade da intervenção, isto é, orientam as práticas dos atores nas fases de conceção, planeamento, implementação e execução, estando também articuladas na forma de encadeamento causal; (iii) a reconstrução da TdM visa providenciar um mapa da lógica processual da intervenção, baseada no quadro teórico-conceptual a partir do qual foi desenhada e implementada a política; (iv) finalmente, as políticas públicas para serem avaliadas nos termos em que foram desenhadas e executadas têm que reconstruir a lógica racional que esteve na sua origem, conectando os inputs da política (as ações ou atividades realizadas) com os outputs desejados (os efeitos ou impactos gerados). A avaliação traduz-se no teste empírico das teorias e hipóteses que estiveram na base do desenho da intervenção e do modelo de operacionalização da mesma.

Relativamente ao segundo objetivo, no âmbito do MAPP-CIES, propomos operacionalizar a TdM em todas as fases que perfazem a vida de uma política pública. Nesse sentido, a TdM diz respeito a todo o ciclo de vida da intervenção, incluindo assim as fases de conceção, de planeamento, de implementação e os resultados e impactos previstos. Assim, propomos a sua operacionalização em três níveis de elaboração, a saber: (i) a teoria programática, incide sobre o desenho global da intervenção; (ii) a teoria da implementação, que incide sobre o planeamento ou operacionalização da intervenção; e (iii) a teoria da produção de efeitos, que incide sobre a execução das ações ou a aplicação das medidas.

O desenvolvimento teórico-metodológico da AIC no âmbito do Modelo MAPP-CIES foi orientado por dois objetivos principais: (i) clarificar ao estatuto da Análise de Impacto Contrafactual (AIC) no processo de avaliação; (ii) potenciar a utilidade do método contrafactual no processo global de avaliação.

Relativamente ao primeiro objetivo importa referir que nem sempre o objeto em estudo permite a sua execução, como foi, aliás, o caso dos CLDS. Quer pela inexistência de dados, quer pela complexidade dos programas e pelo tipo de ações desenvolvidas, efetivamente, a sua prescrição em matéria de modelos de avaliação deve ser criteriosa. Na medida do possível, a execução deste tipo de exercício deve ficar condicionada à existência de

elementos que o possibilitem. Mais ainda, a principal dificuldade que se coloca com a AIC ao nível da testagem dos mecanismos de produção de efeitos nem é a presença ou ausência de dados, mas a limitação desta abordagem a relações mono-causais simples entre as variáveis, o que não acontece em muitos casos. No entanto, as limitações próprias desta metodologia levam a que sua utilização seja circunscrita a dimensões muito específicas das medidas a serem avaliadas, existindo algum potencial por explorar na tentativa de determinação dos resultados globais líquidos da intervenção.

Do ponto de vista prático, o potencial de operacionalização da AIC está muito dependente, por um lado, da natureza ontológica dos objetos avaliados, e por outro lado, da existência, disponibilização, e qualidade dos dados. Em teoria, o processo de desenho das medidas de política com base nos princípios da TdM preveria a identificação dos indicadores de monitorização e de apuramento dos resultados para a avaliação, por forma a que estes fossem contemplados em todo o processo de implementação e execução das medidas. Na prática, verificam-se várias lacunas nesta área, que impossibilitam ou limitam severamente a aplicação deste método. Igualmente, verificam-se vários constrangimentos na disponibilização atempada dos dados, condicionando a execução global da avaliação e o potencial da AIC para informar os restantes elementos da avaliação, o que sugere a necessidade de pensar em formas mais eficazes de envolvimento dos parceiros na avaliação e na necessidade de recolha informada de dados aquando da implementação e execução dos programas. Posto isto, o estatuto da AIC deve ser flexível e não mandatário.

Assim, relativamente ao segundo objetivo, aquando da viabilidade da realização da AIC, o exercício, tendo bastante complexidade, revela algum potencial de operacionalidade, podendo contribuir para a avaliação da pertinência do programa e para a conexão entre a intervenção e os resultados na determinação dos efeitos e dos impactos. Porém, dadas as limitações inerentes ao método, este deve ser articulado com a avaliação baseada na teoria. Uma vez articulado com a TdM considera-se que algumas dessas limitações são colmatadas, pela introdução de uma orientação substantiva. Essa fundação teórico-metodológica alicerça a construção e as opções técnicas do exercício, apoia a interpretação dos resultados, e potencia a capacidade de integrar os resultados produzidos pela AIC num âmbito mais alargado da avaliação, não só no teste dos resultados e impactos da intervenção, mas também com contributos para a avaliação do Planeamento e Implementação das medidas. Logo, o exercício de AIC deve ser efetuado o mais atempadamente possível, de forma a fornecer informações uteis a outras componentes da avaliação. Por último, a explicitação de todos os constrangimentos no relatório da AIC deve evidenciar as condições e as opções técnico-metodológicas do exercício, demonstrando eventuais limitações dos resultados. Essa informação é útil para a cuidada extrapolação dos resultados, e ou a cuidada replicação da intervenção noutro universo, e contribuiu para tomadas de decisão mais assertiva.

Referências bibliográficas

- Bamberger, M., Carden, F., & Rugh, J. (2009). Alternatives to the Conventional Counterfactual: Summary of Session 713 Think Tank. American Evaluation Association, Orlando.
- Cahill, H. (2018). Evaluation and the Theory of Change. In K. Freebody, M. Balfour, M. Finneran, & M. Anderson (Eds.), *Applied Theatre: Understanding Change. Landscapes: the Arts, Aesthetics, and Education*, vol 22 (pp. 173-186). Berlin: Springer.
- Capucha, L., Almeida, J. F., Pedroso, P., & Silva, J. V. (1996). Metodologias de avaliação: o estado da arte em Portugal. *Sociologia Problemas e Práticas*, 22, 9-27.
- Chen, H.-T. (1990). *Theory-Driven Evaluation: A Comprehensive Perspective*. Newbury Park, California: Sage.
- Chen, H.-T., & Rossi, P. (1983). Evaluating With Sense: The Theory-Driven Approach. *Evaluation Review*, 7(3), 283-302.
- Comissão Europeia (2021). Design and commissioning of counterfactual impact evaluations. A Practical Guidance for ESF Managing Authorities. Comissão Europeia, Luxemburgo. doi:10. 10.2767/02762
- Comissão Europeia (2020b). How to use administrative data for European Social Funds counterfactual impact evaluations. A step-by-step guide for managing authorities. Comissão Europeia, Luxemburgo. doi:10.2767/721497
- Comissão Europeia (2020a). Counterfactual impact evaluation of European Social Fund interventions in practice. Guidance document for Managing Authorities. Comissão Europeia, Luxemburgo. doi:10.2767/55495
- Comissão Europeia (2019). Advanced counterfactual evaluation methods. Guidance document. Comissão Europeia, Luxemburgo. doi:10.2767/464242
- Comissão Europeia (2013). Design and commissioning of counterfactual impact evaluations. A Practical Guidance for ESF Managing Authorities. Comissão Europeia, Luxemburgo. doi:10.2767/94454
- Comissão Europeia (2013). Conceber e encomendar avaliações de impacto contrafactuais. Um guia prático para as autoridades de gestão do FSE. Comissão Europeia, Luxemburgo.
- Comissão Europeia. (2013). EVALSED: The resource for the evaluation of Socio-Economic Development. Bruxelas: Comissão Europeia.
- Comissão Europeia (2013). "Evaluation Sourcebook. Method and Techniques". Comissão Europeia, Luxemburgo.
- Csillag, M., Krekó, J., Scharle, A. (2020). Counterfactual evaluation of youth employment policies. Methodological guide. Youth Employment PartnerSHIP. <https://ibs.org.pl/en/publications/counterfactual-evaluation-of-youth-employment-policies-methodological-guide/>
- Davidson, E. J. (2000). Ascertaining Causality in Theory-Based Evaluation. *New Directions in Evaluation*, 87, 17-26.
- Garfield, E. (2004). The Unintended and Unanticipated Consequences of Robert K. Merton. *Social Studies of Science*, 34(6), 845-853.
- Hind, J., (2010). Additionality: a useful way to construct the counterfactual qualitatively? *Evaluation Journal of Australasia*, 10 (1), 28–35.
- Jaffe, P., Crooks, C., & Wolfe, D. (2003). Legal and Policy Responses to Children Exposed to Domestic Violence: The Need to Evaluate Intended and Unintended Consequences. *Clinical Child and Family Psychology Review*, 6(3), 205-213.

- Marôco, J. (2021). *Análise Estatística com o SPSS Statistics*. 8ª edição. Lisboa: Report Number.
- Oliver, K., Lorenc, T., & Tinkler, J. (2020). Evaluating unintended consequences: New insights into solving practical, ethical and political challenges of evaluation. *Evaluation*, 26(1), 61-75.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. Londres: Sage.
- Pearl, J., Mackenzie, D., (2018). *The book of why: the new science of cause and effect*. Basic Books, Nova Iorque.
- Pearl, J., Glymour, M., Jewell, N., (2016). *Causal inference in statistics. A Primer*. Wiley, Reino Unido.
- PO CH (2021). *Aquisição de Serviços de Avaliação do Contributo do PT2020 para a Promoção do Sucesso Educativo, Redução do Abandono Escolar Precoce e Empregabilidade dos Jovens para o POCH. Anexo Metodológico*. Disponível em <https://www.poch.portugal2020.pt/>
- Rogers, P. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1), 29 – 48.
- Rogers, P., Petrosino, A., Huebner, T., & Hacsí, T. (2000). Program Theory Evaluation: Practice, Promise, and Problems. *New Directions for Evaluation*, 87, 5-13.
- Weiss, C. H. (1995). Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In J. Connell, A. Kubisch, L. Schorr, & C. H. Weiss (Eds.), *New Approaches to Evaluating Community Initiatives: Volume 1, Concepts, Methods, and Contexts* (pp. 65-92). Washington, DC: The Aspen Institute.
- Weiss, C. H. (1997). How Can Theory-Based Evaluation Make Greater Headway? *Evaluation Review*, 21(4), 501-524. doi:10.1177/0193841X9702100405
- Weiss, C. H. (2000). Which Links in Which Theories Shall We Evaluate? *New Directions for Evaluation*, 87, 35-45.
- White, H. (2009). *Theory-Based Impact Evaluation: Principles and Practice*. Working Paper 3. International Initiative for Impact Evaluation. https://www.3ieimpact.org/sites/default/files/2017-11/Working_Paper_3.pdf

Anexos

Anexo A: Grelha de análise aos relatórios de avaliação

Nota: os números das questões nas grelhas de análise referem-se ao esquema que questões construídas no início dos estudos de caso para testar o Modelo MAPP-CIES que, na sequência, destes estudos foi atualizado. Ou seja, estes números referem-se a questões que não coincidem com as que foram inseridas no outro documento deste projeto, pois essas resultam do trabalho de análise que aqui se apresenta e da posterior reflexão por parte da equipa sobre o MAPP-CIES.

Fases da avaliação / Dimensões de análise	Objetivos gerais	Q#	Objetivos específicos / Questões a dar resposta
Enquadramento e delimitação do objeto de avaliação	Compreender se a avaliação deixou claro qual o objeto da avaliação; que documentos foram consultados; qual o papel da entidade de acompanhamento da avaliação no desenho e arranque da avaliação; quais são as perguntas de avaliação e quais os indicadores de avaliação; qual o modelo de abordagem metodológica escolhido e as razões pelas opções tomadas.	1	O objeto de avaliação está devidamente identificado e está claro nos produtos de avaliação?
		2 + 2,1	Foram consultados relatórios/documentação dos programas operacionais (PO)/avaliações anteriores? Essa informação é mencionada nos produtos de avaliação?
		3	Os objetivos mais gerias da avaliação são apresentados de forma pertinente e clara?
		4	Foram realizadas reuniões iniciais com os membros das equipas gestoras e de acompanhamento dos PO?
		5	Foi explicitado um modelo/abordagem metodológica à avaliação a ser realizada e para resposta às questões de avaliação (QA)?
Explicitação do problema	Observar se a avaliação procedeu ao diagnóstico, através do levantamento e sistematização da informação (qualitativa e quantitativa), a qual permite identificar os problemas em foco aos quais as medidas visam dar resposta; se existe sistematização sobre as prioridades de intervenção, potencialidades e bloqueios; se foram identificados os problemas reais, potencialidades e riscos aos mesmos; e se foram mobilizados os agentes adequados.	6	Foi realizado um diagnóstico?
		7	Foram identificados os atores-chave, stakeholders, beneficiários das medidas em avaliação, população alvo, etc, no âmbito do diagnóstico traçado?
		8	Foram realizadas entrevistas exploratórias, ou outra formas de auscultação, a atores-chave (stakeholders, entidades promotoras gestoras, beneficiários diretos e indiretos das medidas, comunidade científica e outros elementos da comunidade) que estiveram ou foram implicados/contribuíram para diagnosticar o(s) problema(s) e suas causas?
		9	Foram consultados relatórios, artigos científicos, ou outros documentos relevantes, para a construção/sustentação do diagnóstico?
		10	Recorreu-se a informação/tratamento estatístico?
		11	Identificam-se com clareza quais as prioridades de intervenção, a partir do diagnóstico proposto?
		12	O contexto - ou variáveis determinantes externas (potencialidades e riscos) - e as múltiplas perspetivas dos diferentes atores/agentes foram valorizados no diagnóstico que é proposto?

Conceção	Compreender de que modo a avaliação tratou das etapas que prepararam a implementação da medida no terreno. Assim, pretendia saber-se de que modo foram observados, pela avaliação, os objetivos específicos, as medidas e as metas traçados pela política; os recursos humanos, financeiros e materiais previstos; os atores-chave envolvidos; e o modelo de monitorização e os indicadores de resultado eventualmente estabelecidos. Considerando o papel da TdM, questiona-se se foram observados os fatores que são causalmente encadeados para cada medida para atingir os fins desejados.	13	É estabelecida uma relação com outras políticas no terreno, que concorrem para solucionar os mesmos problemas?
		14	A sinalização das prioridades de intervenção resulta de um trabalho participado entre equipa de avaliação, membros dos PO, stakeholders e/ou outros atores-chave relevantes?
		15	Foram identificados/envolvidos os atores (<i>stakeholders</i> , entidades promotoras gestoras, beneficiários diretos e indiretos das medidas, <i>etc.</i>) que foram implicados/contribuíram para o processo de planeamento do programa/medidas? Quem são estes atores e promotores? Que dificuldades existiram nesta fase? Foi um trabalho participado? Que ações foram desenvolvidas nesse sentido?
		16	Os objetivos identificados para o desenho do modelo, decorrem da sistematização ao trabalho elaborado na fase anterior? Os objetivos da avaliação estão interligados com os problemas que foram trabalhados na fase anterior? Há novos objetivos? Se sim porquê?
		17	Foram auscultados atores (<i>stakeholders</i> , entidades promotoras gestoras, beneficiários diretos e indiretos das medidas, <i>etc.</i>) para o processo de definição dos objetivos? Quem são os atores-chave, stakeholders envolvidos na definição dos objetivos? São os mesmos que estiveram envolvidos no diagnóstico?
		18	A avaliação consegue identificar os objetivos específicos, isto é, mais concretos e mensuráveis?
		19	Estes objetivos desdobram-se em ações concretas com metas associadas? Quais as ações concretas? Quais as metas pretendidas (foram designadas por quem? Pelo promotor da medida, pela EU, outra?)
		20	A relação entre as medidas/ações concretas e os objetivos específicos fica clara nos produtos que resultam da avaliação?
		21	Os principais beneficiários do programa/projeto estão identificados? Como são envolvidos no processo de avaliação? E que papel têm na implementação da medida?
		22	Existem metas de execução intermédia e finais? [Qual o cenário de chegada desejado]
		23	Foram ouvidos atores (<i>stakeholders</i> , entidades promotoras gestoras, beneficiários diretos e indiretos das medidas, <i>etc.</i>) que estão implicados/contribuíram para a definição das metas intermédias e finais?
		24	O programa em avaliação já previa um sistema/ modelo de monitorização e/ou avaliação? Que atores (<i>stakeholders</i> , entidades promotoras gestoras, beneficiários diretos e indiretos das medidas, <i>etc.</i>) foram implicados/contribuíram para o modelo de avaliação e monitorização? A etapa prevista de avaliação foi, ou está a ser cumprida?
		25	Identifica-se uma TdM pertinente e adaptada aos objetivos (gerais e específicos)?
		26	Fica claro que a TdM permite responder às QA?

Anexo B: Grelha de análise da Avaliação do Contributo do PT2020 para o POCH

Fases da avaliação / Dimensões de análise	Q#	Estas questões foram cobertas pela avaliação? (coberta, parcialmente coberta, não coberta, informação desconhecida)	Questão de avaliação concreta que foi colocada, que está subentendida ou que deveria ter sido colocada (n/a para não aplicável)	Notas / descrição
Enquadramento e delimitação do objeto de avaliação	1	Coberta	n/a	Relatório inicial refere: "A presente avaliação inscreve-se no Plano Global de Avaliação do PT2020 e no Plano de Avaliação do POCH e assume a natureza de uma avaliação de impacto com o objetivo geral de identificar em que medida as intervenções do PT 2020 têm contribuído para aumentar o sucesso educativo e, por essa via, reduzir o nível de abandono escolar, em direção à meta inscrita no PNR, bem como potenciar a empregabilidade dos jovens apoiados, face às metas definidas nesta área e tendo em consideração que as diferentes modalidades formativas apoiadas pelos fundos são, em regra, de dupla certificação (escolar e profissional)", (p.3)
	2	Coberta	n/a	O caderno de encargos refere um conjunto de estudos e relatórios sobre o objeto de avaliação. Foram consultadas diversas fontes de informação para compreensão do objeto de avaliação, quer em sede de proposta, devidamente identificadas, quer no relatório inicial. No relatório inicial são também apontadas outras fontes de informação secundária que vão ser mobilizadas de forma a compreender o objeto e a contribuir para a sua avaliação (pp.74-76)
	2.1	Coberta	n/a	idem
	3	Coberta	n/a	Os objetivos gerais apresentados são claros e derivam do caderno de encargos.
	4	Parcialmente coberta	n/a	Sim. Por exemplo, Relatório inicial indica: "a Equipa de Avaliação teve em consideração os comentários do Grupo de Acompanhamento sobre esta componente de análise, efetuados na reunião inicial" (p.1). No entanto, as atas não constam nos Relatórios
	5	Coberta	n/a	Com bastante clareza na proposta técnica e no relatório inicial
Explicitação do problema	6	Coberta	n/a	Foi feito o diagnóstico em dois níveis: UM , pela entidade/político, que consta no Caderno de encargos: Referencia às conclusões do Conselho Europeu de 17/06/2010, e os 5 objetivos da UE para a Nova Estratégia de Emprego e Crescimento / Estratégia Europa 2020; Referencia ao Acordo de Parceria PT2020, ao domínio temático do capital humano e à identificação dos principais "estrangulamentos"; taxas de abandono escolar precoce; E OUTRO , pelo avaliador, que consta no Relatório inicial Há referencia em notas de rodapé à avaliação ex-ante; são apontados dados estatísticos, referenciais europeus e metas a atingir - No relatório não há referência a ter havido um diagnóstico do problema, prévio à política, mas houve uma avaliação ex-ante. - Aquilo que se regista é a construção de uma Teoria da Mudança por parte do Avaliador e, portanto, ex post. A TdM apresentada foca 3 grandes temas, abrangendo um vasto número de indicadores: I. Retenção e abandono escolar, através da evolução das políticas, do quadro de protagonistas

Fases da avaliação / Dimensões de análise	Q#	Estas questões foram cobertas pela avaliação? (coberta, parcialmente coberta, não coberta, informação desconhecida)	Questão de avaliação concreta que foi colocada, que está subentendida ou que deveria ter sido colocada (n/a para não aplicável)	Notas / descrição
				e do contexto de enquadramento do problema II. Abandono escolar precoce: definição do problema, através dos fatores-problema associados aos protagonistas, às políticas e ao contexto; e através das medidas de combate ao AEP (medidas de diversificação das vias educativas e formativas, medidas de prevenção e combate ao insucesso escolar, e medidas de qualificação das infraestruturas, equipamentos e recursos humanos) - Esquema da TM do combate ao AEP na página 24 III. Promoção da empregabilidade: definição do problema através dos fatores-problema associados aos protagonistas, às políticas e ao contexto; e através das medidas de promoção de empregabilidade - Esquema da TM da promoção da empregabilidade dos jovens na página 32. NOTA: de acordo com Relatório inicial, TdM foi construída com base em entrevistas a stakeholders e workshop, ver p.37
	7	Coberta	n/a	Lista de stakeholders a envolver na avaliação na Proposta Técnica, pp. 89-90; e no Relatório inicial, p.37; contribuem para o desenho da TdM. Stakeholders divididos em: Autoridades de Gestão do PO Temático e Regionais; Decisores de política e Administração; Agências públicas com responsabilidades nas intervenções objeto de avaliação; Entidades Beneficiárias; Representantes das Entidades Beneficiárias e dos destinatários finais; e Destinatários finais
	8	Coberta	n/a	idem
	9	Coberta	n/a	O caderno de encargos refere um conjunto de estudos e relatórios sobre o objeto de avaliação. Foram consultadas diversas fontes de informação para compreensão do objeto de avaliação, quer em sede de proposta, devidamente identificadas, quer no relatório inicial. No relatório inicial são também apontadas outras fontes de informação secundária que vão ser mobilizadas de forma a compreender o objeto e a contribuir para a sua avaliação (pp.74-76)
	10	Coberta	n/a	Sim; DGEEC, principalmente
	11	Coberta		Estão identificados 3 aspetos chave nos objetivos gerais - sucesso educativo, abandono escolar precoce e empregabilidade.
	12	Coberta	n/a	De acordo com os guiões de entrevistas propostos no Relatório inicial, e aceitando que o diagnóstico é feito através da construção da TdM, há intenção de valorizar as perspetivas dos atores envolvidos. De igual modo, são pesados fatores de contexto que colocam em risco o sucesso das medidas - ver TdM no Relatório inicial, pp.8-32; ou síntese no Relatório intermédio
	13	Coberta	n/a	Há um esforço de contextualizar os problemas no tempo, com referência a intervenções políticas anteriores àquele objeto da avaliação; há também perspetiva geral de políticas de educação. Ver desde logo a Proposta técnica, pp. 9-15. Ver também TdM no Relatório inicial, pp.8-32; Relatório intermédio pp. 5-10
	14	Coberta	n/a	Ver Proposta Técnica, Envolvimento dos Stakeholders nas diferentes fases do processo de Avaliação, pp. 89-99 + Cronograma nos vários relatórios, que inclui sempre esta participação
Conceção	15	Coberta	n/a	Lista de stakeholders a envolver na avaliação na Proposta Técnica, pp. 89-90; e no Relatório inicial, p.37; contribuem para o desenho da TdM. Stakeholders divididos em: Autoridades de

Fases da avaliação / Dimensões de análise	Q#	Estas questões foram cobertas pela avaliação? (coberta, parcialmente coberta, não coberta, informação desconhecida)	Questão de avaliação concreta que foi colocada, que está subentendida ou que deveria ter sido colocada (n/a para não aplicável)	Notas / descrição
				Gestão do PO Temático e Regionais; Decisores de política e Administração; Agências públicas com responsabilidades nas intervenções objeto de avaliação; Entidades Beneficiárias; Representantes das Entidades Beneficiárias e dos destinatários finais; e Destinatários finais
	16	Coberta	-	Sim, os objetivos do programa decorrem do diagnóstico
	17	Parcialmente coberta	-	Do ponto de vista da avaliação que foi feita, embora pareça ter havido esta preocupação, sobretudo na Fase de Conceção Metodológica (ver Proposta Técnica, p.97), o facto é que os guiões das entrevistas apresentados no Anexo Relatório inicial se centram apenas nas QA. Nota: na avaliação ex-ante é referido: <i>"A participação de diferentes atores públicos e da sociedade civil foi assegurada no processo de elaboração do PO, sendo de destacar a maior articulação entre a tutela governativa pela área da educação e os restantes atores/ parceiros, públicos e privados. No essencial, pretende-se aprofundar o envolvimento dos parceiros ao nível da implementação e acompanhamento dos PO e, por essa via, assegurar a "integração dos princípios do envolvimento dos parceiros". Para o conseguir, está previsto que a Comissão de Acompanhamento do PO, seja uma das sedes privilegiadas para garantir a participação institucional, económica e social, e integre representantes da autoridade de coordenação, dos organismos intermédios e dos principais stakeholders do Programa". p.74</i>
	18	Coberta	-	Apesar de se tratar de uma avaliação de impacto e, como tal, nas QA não haver uma referência clara a objetivos quantitativos, a Avaliação efetuada consegue trazer esta parte mais mensurável para a análise. Além da meta para o AEP, de 10% até 2020, o Relatório Final e os anexos metodológicos contêm mta info quantitativa, sobre metas e realizações. Apesar disto, teria sido bom uma identificação esquemática/quadro síntese do ponto de vista dos objetivos quantitativos de cada ação. Por exemplo, atualmente, no eixo 1 do POCH, estão definidas como metas: 80% de diplomados nas ofertas formativas dirigidas à promoção do sucesso educativo de nível básico (ISCED 2); 85% de alunos transitados para o ano de escolaridade seguinte nos cursos de nível básico (ISCED 2); 70% de diplomados nos cursos de dupla certificação de nível secundário (ISCED 3); 50% de empregabilidade ou prosseguimento de estudos nos seis meses seguintes à conclusão dos Cursos Profissionais
	19	Coberta	-	A TdM identifica, para os problemas a avaliar, um conjunto de medidas, realizações e mecanismos, os quais contribuem para dados processos de mudança e estes, por suz vez, a resultados e impactos. Em alguns casos há metas a atingir, mas na maioria, são apontados e esquematizados objetivos específicos qualitativos. Conforme referido anteriormente, um quadro síntese do ponto de vista dos objetivos quantitativos teria sido bom.
	20	Coberta	n/a	TdM identifica esta relação.
	21	Coberta	n/a	-
	22	Coberta	-	Ver Relatório Final e os anexos metodológicos – info-quantitativa, sobre metas e realizações.

Fases da avaliação / Dimensões de análise	Q#	Estas questões foram cobertas pela avaliação? (coberta, parcialmente coberta, não coberta, informação desconhecida)	Questão de avaliação concreta que foi colocada, que está subentendida ou que deveria ter sido colocada (n/a para não aplicável)	Notas / descrição
	23	Parcialmente coberta	-	Tal como referido anteriormente, do ponto de vista da avaliação que foi feita, embora pareça ter havido também esta preocupação, sobretudo na Fase de Conceção Metodológica (ver Proposta Técnica, p.97), o facto é que os guiões das entrevistas apresentados no Anexo Relatório inicial centram-se apenas nas QA. Nota: na avaliação ex-ante é referido: "A participação de diferentes atores públicos e da sociedade civil foi assegurada no processo de elaboração do PO, sendo de destacar a maior articulação entre a tutela governativa pela área da educação e os restantes atores/ parceiros, públicos e privados. No essencial, pretende-se aprofundar o envolvimento dos parceiros ao nível da implementação e acompanhamento dos PO e, por essa via, assegurar a "integração dos princípios do envolvimento dos parceiros". Para o conseguir, está previsto que a Comissão de Acompanhamento do PO, seja uma das sedes privilegiadas para garantir a participação institucional, económica e social, e integre representantes da autoridade de coordenação, dos organismos intermédios e dos principais stakeholders do Programa". p.74
	24	Coberta	n/a	Plano Global de Avaliação do PT2020 e Plano de Avaliação do POCH
	25	?	-	Solicitar os pareceres dos grupos de acompanhamento
	26	?	-	Solicitar os pareceres dos grupos de acompanhamento; considero que sim, com exceção da questão da eficiência e talvez da sustentabilidade
Implementação	27	Coberta	-	Sim, entrevistas, inquéritos, workshops
	28	Coberta		
	29	Coberta		
	30	Coberta	-	-
	31	Coberta	-	Os vários POs tinham sistemas de monitorização.
	32	Coberta	-	-
	33	Coberta	-	Lista de stakeholders a envolver na avaliação na Proposta Técnica, pp. 89-90; e no Relatório inicial, p.37; contribuem para o desenho da TdM. Stakeholders divididos em: Autoridades de Gestão do PO Temático e Regionais; Decisores de política e Administração; Agências públicas com responsabilidades nas intervenções objeto de avaliação; Entidades Beneficiárias; Representantes das Entidades Beneficiárias e dos destinatários finais; e Destinatários finais
	34	Coberta	-	Os relatórios dão conta de reajustamentos atendendo a alterações de contexto, fruto de monitorização não apenas das medidas, como, sobretudo, das necessidades contextuais e políticas. Exemplo, Relatório Final, secção sobre "Adequação do dimensionamento da alocação de recursos financeiros (globais e relativos) face à ambição de objetivos definidos e às

Fases da avaliação / Dimensões de análise	Q#	Estas questões foram cobertas pela avaliação? (coberta, parcialmente coberta, não coberta, informação desconhecida)	Questão de avaliação concreta que foi colocada, que está subentendida ou que deveria ter sido colocada (n/a para não aplicável)	Notas / descrição
				necessidades dos contextos de intervenção", p.42; + secção "Maior adequação da oferta às necessidades e expectativas dos jovens e famílias", p.57; + "Contínua calibragem da oferta educativa em relação às necessidades do mercado de trabalho", p.60; p.79/80/...
Resultados e Impactos	35	Coberta	-	-
	36	Coberta	-	-
	37	Coberta	-	Por exemplo. através da AIC
	38	Coberta	-	os indicadores estão bem identificados, relacionados com o diagnóstico e TdM
	39	Coberta	-	-
	40	Coberta	-	-
	41	Coberta	-	-
	42	Coberta	-	-
	43	-	-	-
	44	Coberta	-	-

Anexo C: Grelha de análise da Avaliação de Impacto dos CLDS

Fases da avaliação / Dimensões de análise	Q#	Estas questões foram cobertas pela avaliação? (coberta, parcialmente coberta, não coberta, Infor Macao desconhecida)	Questão de avaliação concreta que foi colocada, que está subentendida ou que deveria ter sido colocada (n/a para não aplicável)	Notas / descrição
Enquadramento e delimitação do objeto de avaliação	1	Coberta	-	Está-se a avaliar uma medida, nomeadamente os CLDS, financiados pelo Compete/POPH, pelos PO de Algarve e Lx, e pelo POISE (POR Algarve e POR Lx). + “a) Identificar o grau de eficácia e eficiência e os impactos do Programa no QREN, face aos objetivos propostos; b) Explicitar a forma como as intervenções apoiadas no QREN produziram os efeitos observados, em que circunstâncias e em que contextos o fizeram e os fatores críticos de sucesso (e de insucesso); c) Estabelecer evidência sobre o racional que sustenta o Programa CLDS do QREN; d) Aferir, à luz dessa evidência e dos elementos de contexto regulamentar, socioeconómico e organizacional que enquadram a implementação do PT2020, a adequação dos instrumentos de apoio previstos no atual período de programação face aos objetivos que prosseguem e identificar eventuais necessidades de ajustamento que favoreçam a eficácia, eficiência e sustentabilidades das intervenções.” (Caderno de encargos, p. 21)
	2	Coberta	-	Ver Bibliografia de qualquer relatório. Acresce, por exemplo, secção de Enquadramento do Relatório Inicial, que contem referência a outros POs e legislação + Fontes de informação na exploração das QAs
	2.1	Coberta	-	idem
	3	Coberta	-	Ver Relatório inicial, Objetivos da Avaliação, que cita os objetivos mencionados no Caderno de encargos
	4	Coberta	-	Ver ata no Relatório inicial
	5	Coberta	-	Ver Quadro conceptual, no Relatório inicial
Explicitação do problema	6	Coberta	-	Do ponto de vista do programador: Houve um diagnostico feito pelo ISS, sobre territórios vulneráveis; com base nesse diagnostico, foram criados os critérios para os CLDS + Caderno de Encargos, pg 14: Assumindo-se como um importante instrumento de intervenção de proximidade, em 2007, a conceção do Programa partiu do diagnóstico realizado pela Tipificação das Situações de Exclusão em Portugal Continental (ISS, I.P., 2005), da avaliação das insuficiências detetadas no Programa PROGRIDE (2004-2008) e da necessidade de concretizar de forma mais efetiva as estratégias definidas no Plano Nacional de Ação para a Inclusão (PNAI) (2006-2008). + pg 15: A conceção das ações que operacionalizam os CLDS (nas três vagas) tem em consideração os diagnósticos de base local, desenvolvidos no âmbito das Redes Sociais e valoriza o seu capital de experiência e enraizamento no território, no sentido de mobilizar as redes de parceria imprescindíveis à conceção e operacionalização dos projetos. + CLDS Relatório inicial: para saber que territórios podiam estar abrangidos pelos CLDS, foram

				observadas duas portarias, as quais estabeleciam os critérios. Estes tinham por base: níveis de industrialização, de desemprego, situações críticas de pobreza, particularmente infantil, envelhecimento da população e calamidades (p.15). Do ponto de vista do avaliador: foi desenvolvida uma TdM, que procede a um diagnóstico sobre a relação entre a pobreza e os territórios
	7	Parcialmente coberta	-	Do ponto de vista do programador: informação desconhecida. Do ponto de vista do avaliador: envolveu os stakeholders na reformulação e teste da TdM. Ver Relatório inicial, p.27. Exemplos: PO ISE; SDC; POR Lisboa; POR Algarve; ISS, I.P.; CASES; Associação Nacional de Municípios + Entidades referidas nas Fontes de informação
	8	Parcialmente coberta	-	Do ponto de vista do programador: informação desconhecida. Do ponto de vista do avaliador: idem
	9	Parcialmente coberta	-	Do ponto de vista do programador: informação desconhecida. Do ponto de vista do avaliador: idem
	10	Coberta	-	-
	11	Coberta	-	<ul style="list-style-type: none"> · O Eixo 1 dirigido ao emprego, formação e qualificação; · O Eixo 2 dirigido à intervenção familiar e parental, (preventiva da pobreza infantil – CLDS+ e CLDS 3G); · O Eixo 3 dirigido à capacitação das comunidades e das Instituições; · O Eixo 4, dirigido à Informação e acessibilidade, que existiu apenas na primeira geração de CLDS; · O programa comportou ainda um Eixo de Auxílio e intervenção emergencial às populações inseridas em territórios afetados por calamidades (CLDS 3G).
	12	Parcialmente coberta	-	Do ponto de vista do programador: informação desconhecida. Do ponto de vista do avaliador: revisão da TdM após consulta dos stakeholders
	13	Coberta	-	Sim, sobretudo com outras medidas/eixos de vários POs
	14	Coberta	-	Ver atas de reuniões
Conceção	15	Coberta	-	Sim, o avaliador cobriu esta questão através de entrevistas, estudos de caso e questionários. Guiões dirigidos a Responsáveis políticos pela temática em avaliação, Instituto da Segurança Social, I.P. (ISS, I.P.), Organismo Responsável pela Execução das Políticas Públicas (CLDS, CLDS+)/ Organismo Intermédio sem Subvenção Global (CLDS3G), Autoridade de Gestão do POPH, Autoridade de Gestão do POISE e Autoridades de Gestão do POR
	16	Coberta	-	Relatório intermédio refere: "os contributos recolhidos foram de grande relevância para consolidar e densificar o modelo lógico e a Teoria da Mudança e Abordagem Realista", p.22
	17	Coberta	-	Do ponto de vista do programador: informação desconhecida. Do ponto de vista do avaliador: coberto
	18	Coberta	-	<p>São objetivos específicos:</p> <ul style="list-style-type: none"> · Diminuição do desemprego · Diminuição da pobreza · Inclusão das pessoas em situação de exclusão · Inclusão das pessoas com deficiência e incapacidade · Mais solidariedade intergeracional

	19	Coberta	-	<ul style="list-style-type: none"> · No Eixo Prioritário de Intervenção em Emprego, formação e qualificação, a Melhoria no emprego através de: <ul style="list-style-type: none"> o Melhoria da empregabilidade o Mais procura de emprego (explicitada apenas nos CLDS+) o Aumento do empreendedorismo o Melhoria da inserção profissional dos jovens · No Eixo Prioritário de intervenção familiar e parental, a Mudança na estrutura e organização familiar, através de: <ul style="list-style-type: none"> o Melhoria das competências parentais o Resiliência das famílias aumentada pela capacitação e empoderamento o Mais participação cidadã dos jovens o Menos isolamento de idosos (explícito nos territórios envelhecidos) o Mais solidariedade intergeracional · No Eixo Prioritário de capacitação da comunidade e das instituições, a Mudança socio-territorial, através de: <ul style="list-style-type: none"> o Aperfeiçoamento de processos participativos o Mais auto-organização e Criação de associações · No Eixo Prioritário de Informação e Acessibilidade (até à sua supressão nos CLDS+): <ul style="list-style-type: none"> o Maior acesso às TIC pelos grupos-alvo: crianças, jovens, pessoas em idade ativa e pessoas idosas
	20	Coberta	-	Relatório inicial, p.34: "Independentemente do diagnóstico e do Plano de desenvolvimento social concelhio, os CLDS obedeciam/obedecem [ao] conjunto de eixos pré-determinados. Em cada eixo de intervenção existe um conjunto de ações obrigatórias, existindo a abertura para se desenvolver outras ações, desde que previstas no plano de desenvolvimento social concelhio. Uma vez elaborado o Plano de ação, o mesmo é submetido a parecer do CLAS, posteriormente a ratificação pela Câmara Municipal e, por fim, submetido ao ISS, I.P., enquanto parte integrante da candidatura. Uma vez aprovado o plano de ação, é formalizado o CLDS, mediante a celebração de um contrato entre o ISS, I. P., a câmara municipal ou as câmaras municipais, a entidade coordenadora local da parceria e as entidades locais executoras das ações constantes do plano de ação, onde são definidas as responsabilidades, direitos e obrigações de cada entidade no desenvolvimento do CLDS, bem como os termos e condições do seu financiamento".
	21	Coberta	-	Beneficiários diretos: desempregados, pessoas em transição para vida ativa, jovens, crianças, pensionistas, beneficiários RSI, ... Apesar disto, a medida dos CLDS parece, em última instância, dirigir-se a toda a população do concelho/território abrangido por cada CLDS
	22	Coberta	n/a	O programa dos CLDS tem algumas metas definidas. ver caderno de encargos, pp. 17-18: nº de contratos locais anuais nas várias regiões do país, nº de pessoas a atender/abranger. Apesar disto, há poucos objetivos quantitativos nacionais/gerais, pois cada CLDS define as suas próprias metas.
	23	nao coberta	-	nem o avaliador perguntou ao programador, nem há informação sobre se este os envolveu
	24	Coberta	n/a	Não havia um sistema de monitorização nacional no verdadeiro sentido do termo. Havia informação sobre os vários CLDS em formato papel e/ou dispersa por cada CLDS. Nota, apesar

				disto: De acordo com a Portaria que estabeleceu os CLDS, "Compete ao ISS, I. P., providenciar os instrumentos e os meios que garantam a realização de adequados processos de acompanhamento, controlo e avaliação da execução física e financeira do Programa, podendo recorrer para o efeito à contratação de entidades externas". + Estavam previstas fichas de autoavaliação para os CLDS. Obrigatoriamente, aplicava-se o Plano Global de Avaliação do PT2020 e Planos de Avaliação do POs.
	25	?	-	Solicitar os pareceres dos grupos de acompanhamento
	26	?	-	Solicitar os pareceres dos grupos de acompanhamento; considero que sim, com exceção da questão da eficiência e talvez da sustentabilidade
Resultados e Impactos	27	Coberta	-	Entrevistas, inquéritos, estudos de caso, workshop. Do ponto de vista do avaliador: envolveu os stakeholders na reformulação e teste da TdM. Ver Relatório inicial, p.27. Exemplos: PO ISE; SDC; POR Lisboa; POR Algarve; ISS, I.P.; CASES; Associação Nacional de Municípios + Entidades referidas nas Fontes de informação; Ver também Relatórios Intermédios e Final, com todos os auscultados
	28	Coberta	-	Ver TdM no Relatório Inicial, pp. 29-44
	29	Coberta	-	idem
	30	Coberta	-	-
	31	coberta	-	A avaliação procurou saber isto, embora a resposta seja tendencialmente negativa: "Os projetos foram, regra geral, executados conforme o previsto, embora não exista um sistema de monitorização que, para além da aferição burocrática do cumprimento dos requisitos formais, nos permita ir mais longe na avaliação desta matéria", Relatório Intermédio, p.26
	32	Coberta	-	Questões levantadas nos inquéritos e pedido de informação quantitativa demonstra esta preocupação. Para além disto, Relatório Final dá conta do que foi executado, p. ex, p.37...
	33	Coberta	-	Entrevistas, inquéritos, estudos de caso, workshop. Do ponto de vista do avaliador: envolveu os stakeholders na reformulação e teste da TdM. Ver Relatório inicial, p.27. Exemplos: PO ISE; SDC; POR Lisboa; POR Algarve; ISS, I.P.; CASES; Associação Nacional de Municípios + Entidades referidas nas Fontes de informação; Ver também Relatório Intermédios e Final, com todos auscultados
	34	Parcialmente coberta	-	O avaliador procurou saber isto. Ver, p. ex, Anexos Metodológicos e Guiões: "De que forma e em que medida o Programa se tem ajustado às necessidades dos públicos-alvo?" + "Como avalia o relacionamento com o ISS (enquanto organismo responsável pelo acompanhamento dos CLDS) (Realização de sessões de trabalho; Realização de visitas técnicas aos CLDS; Disponibilização de instrumentos de apoio à monitorização das atividades; Orientações de apoio à implementação das ações; Disponibilização de manuais/guias de apoio; Apoio estratégico ...)", p.218
	35	Coberta	-	Sim, a eficácia é testada de forma extensa através de vários indicadores, coerentes com a TdM e com o restante desenho metodológico proposto pela equipa de avaliação
	36	Coberta	-	A avaliação consegue indicar em que medida o programa foi eficaz (os CLDS estão a realizar os seus objetivos/individuais, mas aquém no impacto nacional; alguns eixos mais bem sucedidos que outros)

	37	Coberta	-	idem
	38	Coberta	-	os indicadores estão bem identificados, relacionados com o diagnostico e TdM
	39	Coberta	-	A avaliação procurou indicar em que medida o programa foi eficiente. Considerou ser relativamente eficiente, mas com limitações e alguma dificuldade em quantificar em que medida o foi atendendo à curta duração dos CLDS.
	40	Coberta	-	idem
	41	Coberta	-	A equipa procurou fazê-lo. Apesar disto, conclui-se que a TdM espelha um racional de mudança que foi apenas parcialmente atingido pelo programa, pois o mesmo teve problemas de operacionalização. O impacto do programa é menor do que se podia desejar do ponto de vista nacional, de diminuição da pobreza, etc.
	42	Coberta	-	-
	43	-	-	-
	44	Coberta	-	-

Anexo D: Guião da entrevista (Coordenadores e Consultores)

O objetivo desta entrevista é falarmos um pouco, e de forma geral, sobre a Avaliação do Contributo do PT2020 para a Promoção do Sucesso Educativo, Redução do Abandono Escolar precoce e Empregabilidade do Jovens através do POCH / Avaliação de Impacto dos Contratos Locais de Desenvolvimento Social (escolher). Reportando a essa avaliação, e enquanto membro da coordenação da avaliação, gostaríamos de colocar algumas questões. Mas antes, gostaríamos de perguntar se autoriza a gravação desta entrevista. Assim, e daquilo que se recorda, naturalmente:

1. Gostávamos de começar pela compreensão do caderno de encargos.

- Eram claros e inequívocos o objeto e os objetivos da avaliação? E a metodologia?
- Caso tenha havido dificuldades na compreensão destes aspetos, de que modo as ultrapassaram?
- Como é que garantiram que a compreensão feita pela equipa de avaliação correspondeu ao que a entidade que solicitou a avaliação tinha em mente?

2. Gostávamos agora de falar um pouco sobre o desenho do modelo de avaliação.

- Como e por quem foram definidas as questões de avaliação?
- Como e por quem foram decididos os métodos e técnicas de recolha e de análise da informação?
- Quais os principais desafios encontrados no desenho da avaliação? Como os superaram?
- Em que medida esse desenho foi ao encontro ou divergiu do solicitado pelo caderno de encargos, e porquê?
- O texto do CLDS continha um diagnóstico, ou um resumo, justificativo dos objetivos delineados pela estrutura de missão do programa?
- Contaram com a entidade contratante para identificação dos stakeholders?

3. Gostávamos agora de centrar a atenção na parte metodológica.

- Se indicou há pouco ter havido um diagnóstico feito pelo programa: Considerando esse diagnóstico, a equipa de avaliação mobilizou o mesmo tal como estava, ou fez uma atualização?
- Se indicou há pouco NÃO ter havido um diagnóstico feito pelo programa: Considerando que esse diagnóstico não foi feito, como é que ultrapassaram essa questão? Foi a equipa de avaliação que fez um diagnóstico referente ao início do programa? Ou fizeram um diagnóstico atualizado?
- Como e por quem foi construída a Teoria da Mudança?
 - Na situação em que foi construída pela equipa de avaliação, qual considera ser o papel da Teoria da Mudança?
 - Que desafios traz construir este racional de “porquê” e “como” num contexto ex-post e por parte de quem vai avaliar, e como os superar?
- Sabemos que a aplicação do Método Contra Factual, solicitada no caderno de encargos, teve algumas limitações. De que forma é que a entidade que solicitou a avaliação reagiu à situação?
 - Pode falar-nos um pouco sobre o que impediu e como ultrapassaram esta situação?
- Como é que correu o envolvimento dos stakeholders?
- No geral, do ponto de vista da execução dos passos metodológicos, que (outras) dificuldades encontraram e como as superaram?

4. Estamos a aproximar-nos do fim. De uma forma geral, como avalia o que foi solicitado, o que foi planeado e o que foi executado?

- Considera ter havido desvios?
- Quais e porquê?
- Como os ultrapassaram?
- Que outros constrangimentos e oportunidades encontraram?

5. Se fosse hoje, o que faria de forma diferente? Porquê e como?

Guião de entrevista a responsáveis nas estruturas dos programas em avaliação

1. Gostávamos de saber um pouco sobre o programa e a primeira questão que gostávamos de colocar é se houve um diagnóstico prévio à criação e implementação do programa?

- Se sim, como foi feito?
- Que atores foram envolvidos nesse diagnóstico?
- Foram identificadas prioridades e definidas medidas nesse sentido?
- Que impactos teve o diagnóstico na criação das várias medidas desenvolvidas?
- Se não, por que motivo?
- Que impactos considera que teve a ausência desse diagnóstico?

2. Foi previsto um sistema de monitorização/implementação do programa/projeto?

- A ser executado por quem?
- Qual foi o seu impacto para a concretização do programa?

3. Quais as expectativas da estrutura de missão do programa, enquanto entidade responsável por este programa/ entidade gestora (Autoridade de Gestão), relativamente à avaliação, ou seja, o que se pretendia com a avaliação dos CLDS/ POCH?

4. Qual foi o papel da estrutura de missão do programa no processo de avaliação dos CLDS/POCH?

5. A estrutura de missão do programa contribuiu para o desenho do caderno de encargos?

- Se sim, qual o seu papel no desenho metodológico?
- De que modo foram equacionadas e decididas as abordagens metodológicas?
- Outras entidades colaboraram para o seu desenho? Quais e de que forma?

6. A avaliação solicitava a construção de uma TdM.

- Qual considera ser o papel da Teoria da Mudança numa avaliação ex-post?
- A que critérios de avaliação era expectável que a TdM respondesse?
- Quais os desafios encontrados na sua construção?

7. Foi solicitado também um exercício de Avaliação de Impacto Contrafactual.

- A que critérios de avaliação era expectável que a AIC respondesse?
- Que desafios encontraram na sua construção?

8. De que forma é que a estrutura de missão do programa articula com as entidades que gerem a execução das medidas financiadas em relação às conclusões e recomendações do estudo de avaliação? Como funciona esta comunicação?

9. De uma forma geral, considera que a avaliação foi capaz de responder às questões de avaliações fundamentais para o programa?

- Quais as questões ou áreas em que a avaliação foi mais bem concretizada?
- Quais as lacunas da avaliação?

10. Se pudesse alterar alguma coisa na avaliação que foi feita, o que alteraria? (se necessário, dar exemplo, na abordagem metodológica, nos métodos e técnicas que foram preconizados, no envolvimento dos stakeholders, na criação de sistemas de informação, etc)

Guião para responsáveis na AD&C

1ª Parte, de carácter geral, sobre a intervenção de AD&C nas avaliações

1.1. Qual o papel de AD&C nos processos avaliativos de programas financiados pelos fundos europeus estruturais e de investimento?

1.2. AD&C costuma contribuir para o desenho dos cadernos de encargos daquelas avaliações? E para as fichas de avaliação inscritas no PGA?

- Se sim, qual o seu papel no desenho metodológico?
- De que modo são equacionadas e decididas as abordagens metodológicas?
- São passíveis de alguma negociação com o avaliador? Se sim, em que situações têm sido feitos ajustamentos? Porquê?

1.3. Considerando as diversas avaliações que AD&C tem acompanhado, com que profundidade é que consideram que as mesmas têm sido capazes de responder às questões de avaliação? Quais os tipos de lacunas ou limitações que tem encontrado? Em que áreas?

1.4. Quais os principais desafios que têm encontrado no que se refere à execução das avaliações dos Programas Operacionais?

- De que forma têm sido ultrapassados esses desafios?
- Que impactos têm tido no desenho de outras avaliações?

1.5. Consideram que as avaliações têm sido capazes de potenciar ao máximo as abordagens metodológicas que as guiaram? Porquê?

1.6. Do ponto de vista metodológico, é habitual ser solicitada uma Teoria de Mudança.

- Na vossa perspetiva, qual é a mais-valia que a abordagem metodológica da Teoria da Mudança tem face a outras abordagens em avaliações ex-post?
- Que critérios têm para validar as TdM, nas várias avaliações que são feitas?
- O que vos parece da utilização da TdM como fio condutor dos procedimentos da avaliação, seja na recolha ou na análise? A TdM tem sido devidamente utilizada para orientar as avaliações?
- E que perspetiva tem sobre a capacidade de aprendizagem dos stakeholders com as avaliações que seguem a TdM?

1.7. Outra abordagem habitual tem sido a Avaliação de Impacto Contrafactual.

- Qual é o papel que a AIC tem tido/deve ter nas avaliações dos Programas Operacionais?
- Que dificuldades têm encontrado na sua construção?
- Como têm sido ultrapassados?
- Parece-vos que a AIC tem sido utilizada devidamente ou poderá ter um papel mais proeminente nas avaliações dos Programas?

2ª Parte, de carácter específico, sobre a intervenção de AD&C na avaliação dos CLDS e POCH

2.1. O que se pretendia com a avaliação dos CLDS? 2.2. E com a avaliação do POCH?

2.3. Quais os principais desafios que encontraram na execução da avaliação dos CLDS? 2.4. E na avaliação do POCH?

2.5. Que impactos é que estas avaliações tiveram, se algum, na intervenção de AD&C nas avaliações que se seguiram?

2.6. Se pudessem alterar alguma coisa nas avaliações naquelas duas avaliações que foram feitas, o que alterariam? (se necessário, dar exemplo, na abordagem metodológica, nos métodos e técnicas que foram preconizados, no envolvimento dos stakeholders, na criação de sistemas de informação, etc).

Anexo D: Relatórios/transcrições das entrevistas

Sobre o caderno de encargos

Sobre os cadernos de encargos no geral, acabam por determinar muito. Acho que no passado não foi tanto assim, mas a tendência é ser de facto mais rígido. É isto que queremos saber e é com estas técnicas. Se a orientação for certa, está tudo bem, até facilita. Mas e se não for? Se não for não dá muita margem para encontrar alternativas. Não há muita margem para negociações. E depois de estar em proposta há sempre esse argumento, “Vocês em proposta disseram que faziam”, e aí já não há volta a dar. Se o caderno de encargos determina muito, uma pessoa ou não aborda o concurso ou aborda. Se aborda, vai ter de dizer que fazemos o que está no caderno de encargos. Se dizemos na proposta que fazemos o que está no caderno de encargos, vamos fazer. Também depende de quem está do outro lado. Existe também um patamar de discussão técnica em que por vezes não está toda a gente no mesmo patamar e às vezes não há um entendimento comum sobre o que se fala. Às vezes tentar explicar porquê que determinada técnica não dá para ali, se do outro lado não houver um entendimento mínimo das especificidades da técnica, não compreende, pode achar que se está só a colocar problemas e que não se quer fazer.

Compreensão do caderno de encargos e negociação metodológica

Em termos de clareza, não houve problemas.

Como cumprir os prazos dos cadernos de encargos?

Diria que antecipar, preparar informação que vai ser necessária para a avaliação, como por exemplo bases de dados, isso podia ser uma coisa a favor. Se os contactos estivessem feitos, ou seja, se tudo o que pudesse ser antecipado fosse antecipado, isso ajudaria. Não sei se era realista, mas na prática, a dada altura estávamos a trabalhar com duas bases: a abordagem contrafactual teve o seu caminho muito autónomo, muito independente, foi trabalhar aqueles dados e analisá-los; não sei se não faz sentido autonomizar dois exercícios, ainda que a determinada altura eles se cruzem... Mas isto porque as condicionantes associadas a esse exercício, como o acesso aos dados, determinam tudo o resto, influenciam o andamento da avaliação. Ou seja, a avaliação no seu todo fica prejudicada.

Sobre a definição das QA

Fora do áudio: Estavam já definidas. O que é possível ao avaliador é sobretudo definir subdimensões.

Sobre Métodos e Técnicas

Ver Caderno de encargos. Ver tópico noutras seções.

Desafios no desenho do modelo de avaliação

Foi um casamento metodológico difícil. Ver tópico noutras seções.

Diagnóstico do programa, havia ou não?

Não havia um diagnóstico propriamente dito, mas havia dados estatísticos que mostravam o que era preciso mudar.

Principais desafios e papel da TdM

A TdM foi construída na primeira fase dos trabalhos. Fizemos entrevistas, houve uma sessão alargada em que apresnetamos uma versão draft, foi discutida e validada com mais stakeholders.

A TdM acaba por ser orientadora do restante desenho metodológico. É no quadro da TdM que se identificam os mecanismos que estão a fazer a diferença, que contribuem para explicar que aquelas atividades contribuem para os resultados; é a fase em que supostamente se devem antecipar os fatores críticos que podem estar a dificultar os resultados a alcançar. Todos os instrumentos são criados para poder testar, para poder confirmar a TdM. Acaba por ser a base para o desenho metodológico. Se quero tentar perceber, por exemplo, se o apoio que foi dado aos Gabinetes de Psicologia e Orientação vocacional tiveram efeito na melhor orientação dos jovens, sei que se calhar vou ter de confirmar em que fases esta orientação é dada, que praticas fazem, qual é a visão que os jovens têm da utilidade daquele processo. Tudo isto tem de ser antecipado na TdM e os instrumentos que crio vão tentar obter informação para testar, para verificar se aconteceu ou não.

Há um desafio de forma. Há a expectativa de que a TdM seja feita seguindo determinadas etapas e produzir um determinado output, capaz de ir ao detalhe e apresentar todas as componentes que podem estar a influenciar o desempenho da política. É um pouco irrealista, como se nós conseguíssemos antecipar tudo, tudo, tudo, todos os fatores que podem ser importantes para compreender como é que aquela intervenção está a produzir aqueles resultados. Acho que esse é um primeiro desafio: pensar que ainda antes de ir ao terreno e recolher informação, que se consegue fazer esse mapeamento completo. Há um esforço para o fazer de forma exaustiva, mas ficará sempre algumas coisas que não se anteciparam. E se não se anteciparam, o desenho dos instrumentos é feito para confirmar aquela “teoria”, e aí pode haver um risco de não nos apercebermos de alguns fatores que podem ser importantes, e não são vistos, porque não estavam previstos à partida, então o manancial de instrumentos não foi suficiente para os apanhar.

O desafio de fazer a TdM ex-post é o de não conseguirmos chegar, por exemplo, àquilo que eram algumas preocupações reais que justificaram a intervenção, e que, se calhar, a posteriori já não se consegue apanhar. Talvez já não se consiga perceber se determinada medida foi criada com uma intenção, uma preocupação qualquer que se não estiver documentada e não conseguirmos chegara ela, recolhê-la, se não conseguirmos chegar junto de quem na altura estava a desenhar a intervenção, ela pode simplesmente não estar na TdM, porque não houve forma de chegar a esse conhecimento. Isso pode significar que não vai ser apanhado pelo exercício de avaliação. A TdM é uma reconstrução.

Existe o risco de enviesamento dos resultados da avaliação, mas não consigo dar nota de uma evidencia que tenha se verificado. Mas o risco está lá, porque na prática estamos a fazer o desenho de uma coisa já com elementos do que ela é hoje, e não o que se pretendia que ela fosse há 3, ou 5 anos. Se aconteceu, em que é que prejudicou ou diminuiu a utilidade do exercício, não sei bem.

Método Contrafactual

Nota: O exercício foi feito, mas com várias dificuldades, sobretudo devido ao objeto da avaliação e à dificuldade de fazer certas comparações. Ver tópico noutras seções.

Implementação

Desafios principais: COVID; algumas mudanças na equipa. O acesso aos dados foi algo crítico, pois não se pode trazer a base de dados e a equipa esteve nas instalações da DGEEC, com todas as dificuldades acrescidas que trouxe.

O cronograma resvalou muito, porque tivemos também a situação do COVID. Mas não foi só o covid: o acesso a dados, a microdados, são sempre processos que implicam estabelecer protocolos institucionais... tudo leva muito tempo. Se isto já estivesse preparado ajudaria, porque quando se começava a trabalhar já havia informação, mas não acontece, porque são também as equipas que têm de ajudar a identificar as variáveis que querem analisar. Portanto, a base de dados é construída já no decurso da avaliação e isso, só por si, justifica que haja uma derrapagem no tempo grande. O cronograma foi largamente ultrapassado. Tudo se reflete depois na rentabilidade do projeto...

Depois os relatórios também são sempre muito discutidos, por um conjunto de entidades. Há sempre estas fases que representam muito tempo; entre entregar um e começar outro vai um mês, porque a equipa defende-se, não quer começar um sem aprovação do outro. Este é um aspeto também muito crítico das avaliações. As equipas também não têm mecanismos de gestão que permitam contabilizar bem as horas, porque se houvesse um mecanismo de controlo do custo-hora, de forma rigorosa, acho que nos assustávamos, provavelmente. Rapidamente o orçamento apresentado na proposta fica desatualizado.

Stakeholders e estudos de caso

Não me recorde de nada muito crítico. Lembro-me de estarmos a fazer a avaliação e começou o COVID e os estudos de caso forma todos feitos à distância. Havia estudos de caso programados para a semana em que fomos todos para casa. Houve uma pausa até percebermos o que ia acontecer. Acho que houve uma adesão. Tínhamos receio que isso desmobilizasse os estudos de caso, nas escolas não fomos logo, logo, porque estavam a tratar das aulas online, mas quando tudo normalizou, retomámos os contactos e acho que correu bem.

Não me lembro de grandes dificuldades. Tivemos entrevistas mais institucionais, mas não me lembro de nada.

Os alunos foram ouvidos no âmbito dos estudos de caso. O diretor da escola, alguns professores, até encarregados de educação em alguns casos. Empresas também onde estagiaram.

A seleção dos estudos de caso foi feita a partir de uma matriz de critérios, com diversidade territorial, com as várias regiões financiadas, tipo de escolas, para incluir escolas privadas, públicas, indicadores escolares, como taxas de abandono, casos de sucesso e insucesso, entre outros, mas sem dificuldades

Resultados da avaliação/ Balanço

Não encontro grandes desencontros entre o planeado e o executado. Há sempre ajustamentos, aquando do confronto com os dados, mas diria que as técnicas que estavam previstas foram aplicadas, os stakeholders foram ouvidos, foram produzidos os relatórios predefinidos em caderno de encargos... demoraram talvez mais tempo. (...). Agora lembrei-me. Acho que em proposta nós tivemos logo algumas reservas quanto a fazer o Contrafactual, então penso que a proposta não é muito assertiva nesse aspeto. Quisemos deixar em aberto a possibilidade de ser objeto de alguma discussão. Mas como entre a proposta e o relatório inicial ficou definido que a era para aplicar e tínhamos de encontrar um desenho para o fazer, a partir daí foi encontrar forma de o concretizar. Entre a nossa proposta e o relatório inicial houve, efetivamente, um ajustamento. Acho que na proposta nem chamámos Análise Contrafactual, usámos outro nome, para nos defendermos um pouco. Propúnhamos uma comparação, mas que não era Análise Contrafactual pura. Foi nisso que tivemos de atalhar caminho, porque havia essa intenção e tivemos de encaixar isso no desenho metodológico.

Hoje, o que faria diferente?

[Do ponto de vista metodológico] houve sempre uma consciência de que estávamos a produzir um exercício com muitas reservas em relação à utilidade e à leitura que se podia fazer dos resultados do exercício. Podíamos ter tentado uma posição mais afirmativa. Mas houve uma altura em que ficou claro que não havia margem. Era uma negociação e talvez nessa fase pudéssemos ter sido menos flexíveis. Mas se tivéssemos sido assim, talvez tivesse havido uma rutura do contrato e a avaliação não ter avançado. Tivemos de ponderar tudo isto, na altura. Por outro lado, por termos feito o exercício, pode ter prejudicado a rentabilidade do projeto, pode ser sido uma peça que podia ter sido desenvolvida de outra forma, (...), mas foi muito valorizado. Quando se comunica os resultados essa parte é muito importante. Os resultados que decorreram dali são muitas vezes o que passa para fora. [A parte quantitativa, de resultados em percentagens] é muito valorizado. Até dá visibilidade ao exercício. Prejudicou na gestão do projeto em si, nos recursos que tivemos de afetar.

No PO CH, a avaliação está completa, está tudo fundamentado. Se tivéssemos mais liberdade, talvez tivéssemos feito outro desenho metodológico. Era interessante assim, depois, confrontar os resultados e perceber.

Sobre o caderno de encargos

O caderno de encargos do CLDS foi o primeiro modelo de avaliação do PT 2020. Era um caderno de encargos modelo das avaliações que se seguiram do PT 2020. Foi das primeiras avaliações temáticas.

Sobre os objetivos: estavam claros, mas eram muito ambiciosos face ao orçamento e às exigências do grupo de acompanhamento. Os objetivos eram claros, as QA também; os critérios também, que já incluíam o Valor Acrescentado Europeu.

Havia uma abertura metodológica muito grande, pediam ABT, mas não se prescrevia a TdM ou outra abordagem. Optaram pela QCA/Análise Comparativa de Casos, em conjunto com TdM, que presidiam ao exercício. Depois aplicaram várias técnicas, como estudos de caso, questionários, entrevistas aos stakeholders mais relevantes, workshops, etc. Centraram muita atenção na avaliação estudos de caso.

Principais aspetos problemáticos mencionados:

1. Custo da avaliação; orçamento muito pequeno e desastroso. A sustentabilidade financeira do projeto estava pouco assegurada, mas concorreram.
2. Havia um problema com o caderno de encargos, que era avaliar as edições anteriores, que já tinha ocorrido há 10 anos. Os técnicos e staff de projetos da 1ª e 2ª edição do CLDS já não estavam presentes. Esta parte foi muito difícil, avaliar uma coisa que já passou há muitos anos. E o caderno de encargos pedia isto e foi difícil, avaliar o QREN e o PT2020.

O caderno de encargos tinha alguns problemas generalistas, como no caso da TdM. Como fazer a TdM? Agora já são mais prescritivos, já estão mais fechados do ponto de vista da metodologia, mas nos primeiros anos havia ainda uma abertura e uma ambição relativamente desmedida e não muito clara no caderno de encargos. Refere: “nós estamos agora a realizar um processo de avaliação onde tem havido problemas nessa prescrição. Existem opiniões e crenças metodológicas diversificadas. Do outro lado está um aparelho burocrático que olha para o caderno de encargos que diz que a TdM tem de ter realizações, resultados, etc, e que define a cadeia de mecanismos que mediam as relações aos resultados, e isto nunca tem sido consensual”.

A TdM nunca é consensual. Não há consenso na prescrição. Hoje em dia, além da prescrição da constituição da TdM, há metodologias quase obrigatórias, como a análise contrafactual, em certas avaliações; ou entrevistas aprofundadas, por exemplo, e sem TdM. Há outras de análise macroeconómica, com prescrições muito específicas. Mas a maior parte dos cadernos de encargos são menos prescritivos, ou seja, só querem a TdM e não dizem quais são os passos; internamente as equipas definem. Tem havido uma evolução nos cadernos de encargos, com lições de aprendizagem que vem das avaliações. Mas ainda existem aspetos muito problemáticos, por exemplo, como fazer uma proposta para responder ao caderno de encargos.

Compreensão do caderno de encargos e negociação metodológica

O contratante, nomeadamente os programas operacionais, pontuam as propostas e fazem comentários. O que têm pedido é que as equipas vencedoras alterem alguns aspetos que estão nos relatórios finais de avaliação das propostas. Na altura isto não foi feito, porque não havia conhecimento do POISE sobre estas matérias.

Considera que foi uma negociação pacífica, talvez pela presença de LC e PP, que tinham muita experiência e sido professores de muitos do grupo – talvez tenha havido um enviesamento por isso – “são os mecanismos escondidos da realidade, nem sempre previstos”.

Mas tivemos alguns problemas. Por exemplo, a explicitação do exercício do QCA, porque não havia muita prática, nem mesmo internacionalmente. Era um conjunto de pressupostos técnicos que nunca tinham sido testados no terreno, nem em Fundos Estruturais. Isto deixou o outro lado desconfiado, sobre a realização by-the-book do exercício.

Também problemas de formulação do modelo de produção de efeitos. Já não na TdM, que era muito vaga, atendendo também à natureza do objeto, mas no desenho das hipóteses, na desagregação que se queria – já se exigia um anexo metodológico onde tudo estivesse mapeado, um mar de coisas onde têm de ser cruzadas as questões dos questionários, dos estudos de caso, vão responder a que indicadores, para que questão, utilizando que método... “Esta componente demasiado burocrática, que tem sido requerida nas avaliações atuais, acho que nasce nesta avaliação dos CLDS, foi quase pioneiro em orientar a ADC”.

Como cumprir os prazos dos cadernos de encargos?

Haver dados sistematizados, a TdM já desenhada... O incumprimento de prazos deve-se ao momento inicial, o tempo entre a assinatura do contrato e a entrega do relatório inicial. Esta mal planeado, é irrealista e atrasa todo o processo, porque se a aprovação daquele relatório não há trabalho de campo, nem teste da TdM.

Sobre a definição das QA

O avaliador tem capacidade para criar subquestões associadas às questões; em alguns casos há reformulações, mas nem sempre há abertura para isso, até porque a Comissão Europeia obriga aos habituais critérios – coerência, relevância, eficácia, eficiência e impacto, mais recentemente valor acrescentado europeu. As questões ao obrigatoriamente definidas no caderno de encargos, mas nas primeiras reuniões o avaliador consegue compreendê-las melhor e sugerir que seja, por exemplo, desagregadas em subquestões, porque é mais fácil para dar resposta; ou criar um foco maior em dado aspeto. Tem havido liberdade para os avaliadores, já em fase de proposta, reformularem ou criarem novas questões. Há prescrição das principais questões, mas o avaliador tem alguma liberdade para as desarrumar.

Sobre Métodos e Técnicas

Nos CLDS puderam definir as abordagens metodológicas mais micro, de técnicas a mobilizar. Mas há alguma rigidez, cada vez mais, porque os Gabinetes Jurídicos fecham o âmbito dos cadernos de encargos, então é preciso cumprir o mesmo, ou a proposta feita. Atualmente é um processo muito burocrático por causa disto. Há às vezes aspetos sugeridos em proposta, mas que depois se vê que não faz sentido, que não é útil e há algumas dificuldades em negociar não fazer isso, para fazer outra coisa.

Desafios no desenho do modelo de avaliação

O orçamento foi uma decisão tomada, era baixo. Propuseram um baixo número de casos, porque era a parte mais onerosa; e acabaram por fazer ainda menos, mas muito bem fundamentado e com um certo trade-off; tiveram de fazer outras análises, para compensar o deixar cair de alguns casos.

Um grande desafio foi a inexistência de informação sistematizada sobre os casos, sobre os projetos. Foi uma avaliação baseada nas evidências recolhidas no terreno, nos questionários, estudos de caso, na QCA; foram as principais dificuldades. Depois também alguns pareceres muito chatos, que vêm com a aprovação dos relatórios. Mas correu bem, no geral, com uma boa taxa de resposta, com a participação de muitas pessoas – claro que isto não foi tao fácil para as primeiras gerações dos CLDS.

Se houvesse informação sistematizada, informatizada, talvez se pudesse ter feito outras coisas, mas não havendo, era impossível ir para um armazém analisar centenas de dossiers. Nem na versão 3G havia isto, quanto mais na CLDS+, ou na primeira, CLDS.

O que tentaram acautelar noutras avaliações era dar ao avaliador antes, ou lançar avaliação só quando houvesse uma base de informação sólida dos projetos. ADC teve essa postura, mas nunca houve nenhuma avaliação com esse acervo de informação muito robusto.

Diagnóstico do programa, havia ou não?

O diagnóstico que havia era global, muito centrado. Nos grandes objetivos dos CLDS, como o desenvolvimento social dos territórios, o combate à pobreza, sobretudo estes dois grandes impactos. Havia teoricamente uma base onde eram selecionados os CLDS. O diagnóstico base do programa e a seleção dos territórios era feito com base num trabalho já antigo, chamado “Situações de Exclusão Social em Portugal”, algo do género, uma publicação do ISS, onde há uma tipificação dos territórios, com algumas características e, portanto, era a partir daí. Isto explicava a seleção dos CLDS. Recuperaram esta publicação e aquilo que concluíram, já no final da avaliação, é que esta tipologia não é muito operacional. Também não nunca houve um diagnóstico mais fino. A articulação era com os diagnósticos das redes sociais, mas como o programa tinha um conjunto de eixos mais rígido, nem sempre casava esse diagnóstico social dos territórios com o seu plano de desenvolvimento social, considerando esses eixos obrigatórios dos CLDS. Havia muitas ações obrigatórias e nem sempre havia essa necessidade nos territórios.

Foi a equipa de avaliação que fez o diagnóstico aquando da construção da TdM. Foi fácil, atendendo à equipa que tinha pessoas que trabalhavam estas áreas, pelo menos na parte global.

Principais desafios e papel da TdM

Nota: Foi a equipa de avaliação que construiu a TdM.

A apropriação da TdM para este tipo de exercícios é problemática, quando se cruza com as respostas de avaliação, que são essas que temos de responder. Portanto, como cruzar toda a análise do teste da TdM com responder a questões de coerência, impacto, eficiência? A TdM costuma ficar só associada à eficácia. Esta foi a estratégia que usaram, de impacto também, mas muito mais na eficácia.

Nesta avaliação, ADC queria fazer by-the-book e impor princípios ainda pouco consolidados. Os guias de avaliação E. Valsa, que presidem à avaliação de Fundos Estruturais, vêm de projetos diferentes. Foram desenhados para projetos de cooperação internacional, pagos pelo Banco Mundial, Banco Africano, etc, onde há um diagnóstico muito marcado, onde se quer mudar uma realidade que também está muito circunscrita, e onde é fácil desenhar aquilo que é a mudança desejada. No caso da avaliação de Fundos Estruturais é muito mais complexo fazer ABT, por estas razões. Nós temos um grande output a resposta a questões de avaliação específicas e não um exercício metodológico de teste, de teste de hipótese. Nos exercícios mais atuais utilizam-se mais os mecanismos.

O caminho até hoje tem sido tortuoso para estabilizar o modelo de TdM e não há um só caminho. Não há consenso entre avaliadores e constantes. Há alguma rigidez e burocracia. Neste contexto, desenharam um modelo muito vago da TdM, muito afastado daquilo que é requerido hoje. Todo o modelo de produção de efeitos que vai desde os problemas, atividades, realizações, outputs e resultados; depois há quem distinga resultados de primeiro nível, de segundo nível; ou intermédios e finais; e depois impactos. Foi muito difícil porque se tratava de um programa que continha 4 eixos, mas eixos muito abertos, com atividades muito distintas e era difícil mapear os resultados previstos para essas atividades. Ainda assim fizeram-nos. Testaram quase 20 ou mais hipóteses. Foi dos primeiros exercícios do PGA (Plano Global de Avaliação) já com a TdM a presidir o desenho metodológico da pesquisa e a influenciar os seus resultados.

Não devia ser pedido aos avaliadores que fizessem a TdM, a não ser que se contratasse esse trabalho à parte. Quem devia construir a TdM devia ser o programador, numa fase ex-ante. E as avaliações ex-ante deviam servir para isso, mas não é a prática. O diagnóstico também tem de estar realizado por quem programa a medida, por quem a planeia. Ser os avaliadores a recuperar o racional de uma ação do passado leva a desajustamentos. Até porque o formato das teorias exigido é muito ambicioso. Os relatórios iniciais do PGA estão a durar muitos meses, se não 1 ano, o que é contra os princípios do caderno de encargos, onde devia ser feito num mês. Na avaliação dos CLDS o caderno de encargos não foi muito realista quanto ao relatório inicial e a inclusão da TdM; e continuam a não ser. Exige-se um relatório metodológico um mês passado o assinar do contrato, com o desenho da narrativa da mudança, que muitas vezes não é explícita, nunca foi, foi feita sem grande suporte de evidência no terreno e anda a reboque das oportunidades dos fundos de financiamento europeus. Arranja-se um diagnóstico muito global para enquadrar as oportunidades de financiamento. Quando chega à componente de avaliação, nós é que sofremos com isso.

É com grande dificuldade que ultrapassam estes constrangimentos relativos ao desenho da TdM numa avaliação ex-post. Cumprir os tempos é quase impossível. A TdM devia estar previamente realizada e nós devíamos testá-la, através de metodologias. Esse devia ser o trabalho da avaliação. Nós estamos a recuperar as lógicas do planeamento e a fazer a avaliação, ou seja, está a ser requerido um conjunto de componentes que deviam já estar feitas. Temos defendido que deve haver financiamento para haver equipas a fazerem TdM de políticas que serão avaliadas daqui a algum tempo, mas nunca foi acolhida.

[Fala sobre a criação da PlanApp].

Refere que a ausência de diagnóstico e de TdM não é um caso português; é um caso europeu. Consultaram repositórios de avaliações e perceberam que só eles estavam a fazer avaliações com tal exigência... Diz que em Portugal fazemos mais com menos.

Fazer a TdM por quem avalia pode enviesar os resultados da avaliação. Há aqui uma ideia de uma certa engenharia social associada à avaliação, atualmente. O que leva os beneficiários a alcançar resultados, nem sempre são fatores racionais, explícitos. Estar a prever um modelo fechado, que depois vai ser testado, exige muito mais tempo e pessoas, e não se coaduna

muito com o que é pedido nos cadernos de encargos. Desenha-se uma TdM, mas depois o teste empírico da mesma pode levar a um exercício estéril, porque o que está previsto na TdM não acontece, ou há outras coisas fundamentais que levam a que os resultados aconteçam. Este é um dos principais desafios sobre as ABT. E mesmo com TdM previamente realizadas, a realidade é complexa. Há situações em que a TdM tem de ser ajustada, introduzidos aspetos que entretanto foram encontrados na base empírica, que explicam os resultados, e não a base teórica inicial. Pode-se começar com uma TdM inicial e acabar com outra no final, havendo neste esforço grandes perdas. Há uma grande fragilidade de ABT, por este conjunto de fatores aqui falados, e pela pouca adequação à avaliação de FEE, onde a Comissão Europeia quer saber a resposta a critérios muito específicos, como a relevância, a coerência, a eficácia, o impacto. Diria que a TdM, tal como está perspectivada nos cadernos de encargos, ela é apenas um auxílio à resposta à eficácia. Tudo o resto acaba por ser mais lateral. Até porque depois a TdM não é recuperada para nada, tão pouco para reparar outros quadros de financiamento/programação.

Método Contrafactual

Seria descabido fazer um exercício de comparação de casos no contexto do CLDS. Não havia bases de dados, não havia informação sobre qualquer grupo de controlo. Quer o ISS, quer o POISE, compreenderam que não era possível.

As especificações técnicas nos cadernos de encargos têm de mudar. Não sei se se trata de irrealismo por parte de quem constrói o caderno de encargos pedir exercícios como o Contrafactual, que não são possíveis de realizar em certas avaliações, como foi a dos CLDS. (...) Teoricamente podia ter sido realizado, mas na prática não havia condições para isso.

Implementação

Não foi uma avaliação fácil, do ponto de vista da sua implementação. (...)

Problemas também na implementação, com recolha de informação, acesso aos projetos que foram usados, a própria sistematização de informação que esses projetos não tinham – as equipas não estavam nada orientadas, e continuam a não estar, na 4G, para a recolha e monitorização estratégica de dados que vão servir para os exercícios de avaliação futuros.

Stakeholders e estudos de caso

A componente dos stakeholders também vem do modelo de caderno de encargos do PGA, que já tem de estar altamente definida nas propostas. Portanto, a identificação já faz parte do caderno de encargos. Aí o trabalho estava mais facilitado. No Grupo de Acompanhamento estava a entidade central, ou seja, quem executava o programa, que era o ISS, tínhamos a gestora do departamento de programas, que tinha uma postura muito defensiva. Fazer uma avaliação nestas temáticas, com estes stakeholders, deixa sempre as pessoas um pouco à defesa, na discussão dos relatórios, nos aspetos negativos da implementação do projeto, naquilo que é defendido... “Eu diria que o trabalho foi quase todo nosso, porque o ISS não tinha informação física de caracterização dos projetos”. Tinha em dossiers de papel num armazém a 30km de Lisboa e era impensável ficar um mês num armazém a fazer esse tipo de análise. Por tanto, não contaram com nenhum sistema de informação. O POISE tinha informação muito vaga, insuficiente para poderem selecionar casos. Então o que fizeram foi pedir à direções distritais da Segurança Social que seleccionassem os melhores e piores projetos, com base num conjunto de critérios definidos pela equipa. Nem todos responderam. Houve um workshop que também permitiu ver um pouco disso e contribui para a seleção dos casos. Tentaram não se responsabilizar pela seleção dos casos, ou seja, que fossem as próprias direções regionais a selecionarem os projetos. Eram critérios que já tinham identificado que eram necessários para a QCA. Queriam por o ónus nas direções regionais, porque não havia nenhum sistema de informação que permitisse selecionar os casos.

O envolvimento com os stakeholders foi muito fácil. Refere que ao tratar-se de um programa desta natureza, os atores que o mobilizam e a entidade coordenadora – que não tem nenhuma reflexão sobre o seu trabalho – acaba por achar este exercício de avaliação muito importante. Tiveram respostas positivas e abertura grande para visitar os territórios e estar com as pessoas, nos estudos de caso. Contaram bastante com o IEFP, equipas técnicas (foi fácil obter informação, porque têm muito conhecimento da área) e beneficiários (identificados pelas próprias equipas técnicas dos CLDS, ou IEFP – esta foi a componente mais difícil, porque ou não apareciam, ou não sabiam que tinham participado naquele projeto, ou não conseguiam elaborar sobre isso; foi a parte mais difícil), por vezes CPCJs e outras entidades com assento nos CLAS, além das Câmaras. (...)

Resultados da avaliação

Os técnicos dos CD ainda hoje sentem dificuldade em ter um Manual de Avaliação, algo que foi proposto no final da avaliação. Este Manual, ou Guia de Avaliação de Indicadores, para saber como medir os resultados dos projetos, que foi proposto no final da avaliação, ainda não foi conseguido. Foi lançada a 4ª Geração quando ainda decorria esta avaliação e sem ter em conta os resultados da avaliação.

Balanço

Tirando todas as atribuições metodológicas, foi mais ou menos linear. Algumas metodologias em que havia a expectativa que se fizesse, como o Contrafactual, foi talvez o único aspeto não concretizado. Mas no geral, considera que realizaram a proposta técnica e deram resposta ao caderno de encargos. Refere que a negociação com o Grupo de Acompanhamento nem sempre foi fácil, também na foi fácil lidar com a inexistência de sistemas de informação, em ter os beneficiários presentes nos estudos de caso. Mas a parte metodológica em boa parte foi cumprida. Não houve grandes desvios. Houve pequenos desvios na concretização metodológica, que foram os que já falámos. Mas conseguiram realizar praticamente tudo o que se propuseram. Na gestão do tempo é que foi aquém – mas nunca nenhuma avaliação cumpre conforme estão os cadernos de encargos. Já sinalizaram isto a ADC, mas não mudou grande coisa. O que não é linear, nem correu tao bem, foi o follow-up das recomendações. Tivemos acesso a um documento comentado pelo ISS, porque a maior parte das recomendações eram para eles. Nada foi implementado e a justificação era sempre a não exequibilidade, ou a recomendação não lhes ser dirigida, mas antes para o POISE, depois o POISE a dizer que não, que é do ISS... este processo final que de certo modo é o mais importante, não tem sido conseguido, a apropriação das recomendações pelos programas. No caso do CLDS este aspeto foi pior, porque quem desenhava a versão 4G nem nos contactou para saber a que tipo de conclusões já tínhamos chagado, muito menos agarrar na avaliação e ver os resultados, procurando desenhar o programa de acordo com as recomendações. Foi construído um novo programa, mas com alterações muito ténues. Não é que haja obrigatoriedade em acolher as recomendações, mas neste caso não foi nenhuma. E de notar que quem estava a desenhar o novo programa, era quem estava envolvido nesta avaliação.

Hoje, o que faria diferente?

Teria investido mais na estruturação do QCA e no aprofundamento de certos fatores. Podíamos também ter confirmado logo nas primeiras reuniões se havia certa informação, sugerir que seria precisa e falar logo sobre isso. Partimos do princípio que havia informação, mas não havia, pelo que a avaliação foi baseada na recolha da informação da equipa, e nos estudos de caso.

Pensando nos objetivos da avaliação do POCH e nos objetivos da avaliação dos CLDS, em que medida consideram que os mesmos eram razoáveis? (pergunta propositadamente aberta; pergunta seguinte estreita o âmbito)

P1: Nos CLDS, o problema foi o período que se queria avaliado, porque queriam que se avaliasse todas as gerações dos CLDS, o que trouxe alguns problemas de conseguir contactar os beneficiários e equipas de gerações anteriores. Este foi o objetivo que gerou mais problemas.

P2: considera que não tiveram dificuldade, nem mesmo na redação da proposta. Houve uma dúvida, que era perceber se queriam uma avaliação de impacto, ou no caso dos CLDS3G se era ongoing. Mas nada que fosse um desafio maior.

Considerando a janela de tempo e o orçamento que tinham, em que medida os objetivos eram exequíveis?

P2: Os CLDS era uma avaliação de 1 ano, o volume financeiro - acho que cresceu a dotação orçamental nas últimas avaliações – mas penso que numa perspetiva geral, as avaliações tem tido uma dotação orçamental mais reduzida, comparando com outros quadros comunitários. A dotação orçamental não é significativa, sobretudo tendo em conta que se pede um conjunto de instrumentos de recolha de informação que pressupõem deslocações, inquéritos, tratamento de bases de dados, etc, com um caderno de encargos que pede um perfil de consultores com um perfil que não é económico. As pessoas têm de ter uma experiência profissional muito grande e específica, na área em concreto, pessoas com Phd ou mestre... O seja, é difícil cumprir

os requisitos a não ser que se recorra a mão-de-obra robusta e adequada. Tendo em conta que era um projeto de 1 ano, os CLDS tinham um conjunto de áreas de intervenção muito grande – educação, emprego, violência doméstica, etc. – procurámos pessoas com conhecimento nestas áreas e em avaliação. Tendo em conta isso e o período de 1 ano, mais os instrumentos e o trabalho de campo em si, não é bem remunerado.

P1: Para além de que os CLDS demoraram muito mais tempo do que foi previsto. É a exigência do trabalho que é pedido, que não se comporta nem com o tempo que está definido, nem com o orçamento.

Como resolveram estas dificuldades?

P2: nos tivemos um curso de gestão de projeto em que ensinaram que há um triângulo: o tempo, o custo e o âmbito. Se uma das coisas é diminuída, as outras têm de compensar. Se o tempo é curto, temos de investir mais em RH. Como o dinheiro não aumenta... o alargamento dos prazos nem sempre é nossa responsabilidade. No caso dos CLDS houve mta dificuldade de acesso a informação e em articular com os vários stakeholders para recolha de informação. Então o tempo é prolongado não só por “responsabilidade” da equipa de avaliação.

Depois, os relatórios iniciais nunca são entregues na data certa, sobretudo quando há entrega de TdM. Na altura dos CLDS não era compagável fazer a TdM e entregar o relatório inicial no espaço de tempo que é dado, que era 1 mês.

Sobre o tempo; quais as implicações disto na TdM? As equipas são montadas no projeto, mas são uma projeção; há um grupo de pessoas que estão garantidas, mas outras que não. O processo de montagem da equipa é mtas vezes superior a 1 mês, que é o tempo de entregar o relatório inicial. Do lado da encomenda já há uma visão clara do que se quer em termos de TdM?

P2: do lado da encomenda, penso que há uma visão clara daquilo que quer. Até porque isso depois sente-se quando se está a discutir a pp TdM. Mas não esquecer que a TdM nunca é explícita, é sempre construída pela equipa de avaliação. Não há uma TdM a priori que vamos avaliar. Temos de ser nós a construí-la, e que construída e desenvolvida com quem tem know-how na área. Daí a importância do recrutamento de especialistas. Nos CLDS tivemos 1 mês para apresentar a TdM, que não está desenhada a priori, teve de ser construída e desenvolvida, e, simultaneamente, todos os instrumentos de recolha de informação, os quais também, supostamente, têm de ser desenvolvidos para responder à TdM que a equipa de avaliação desenha, e que vai ser negociada e que tem depois implicações nos guiões e afins. Temos desenvolvido uma estratégia, porque 1 mês não chegava para tudo, apresentamos a TdM e os instrumentos de recolha de informação vão posteriormente, depois de estabilizada a TdM, para evitar um investimento na definição e programação de instrumentos que depois tem de ser alterado.

A ideia é ter pessoas com know-how que lhes permita um exercício rápido no que respeita à TdM, que sabe quais são os mecanismos de mudança que têm de ser ativados. As propostas são apresentadas num prazo, com equipas contactadas para um determinado efeito e, muitas vezes, o hiato entre a apresentação da proposta e a adjudicação, tb tem os seus percalços, nem sempre temos as pessoas pensadas inicialmente. E nem todas as pessoas com know-how da área conseguem traduzir para um esquema esta questão da TdM, que tb leva o seu tempo.

O processo de montagem de equipa, de contratação, não demora pelo menos 1 mês, o mesmo mês que demora a entrega do relatório inicial?

P2: nos CLDS não tivemos de contratar ninguém, pq o pessoal da equipa era das 3 entidades.

P1: na gestão das equipas, juntar as pessoas para trabalhar, nunca foi um processo que levasse a atrasos nas avaliações. No caso dos CLDS era uma equipa do consórcio e pessoas que já se conhecessem, trabalham juntas e têm uma relação que agiliza o trabalho e a participação de todos no projeto. Há sempre uma reunião inicial, discutem-se as disponibilidades.... A construção

da equipa nunca foi determinante para começar a avaliação. Acho que é mais os conhecimentos das pessoas relacionadas com a teoria e a montagem da TdM. Por isso temos peritos de avaliação, que já fazem isto há mais tempo; e peritos da área, para uns orientarem os outros.

Como é o processo de validação? Parece um processo que não tem prazo... sim ou não? E este processo de validação traz alguma coisa de novo, leva a novas recolhas de informação, novos instrumentos metodológicos, ou estamos a falar do que aquilo é e do que deve lá estar, mas a informação que têm para construir aquela TdM é a mesma que usaram no relatório inicial, só se está a falar de interpretações sobre o desenho final, não sobre lacunas ou processos que se quer desenvolver e que levavam a nova recolha de informação?

P1: é um pequeno pesadelo! É mais uma questão de formato, porque face o que é apresentado em proposta, claro que há todo um trabalho de pesquisa, as entrevistas e tudo o que estiver na metodologia. Nós grande parte das discussões da TdM não exige propriamente uma nova recolha de dados, é mais interpretações, diferentes interpretações sobre o que está na TdM.

Quanto da proposta é que se consegue aproveitar para o relatório inicial? Há um percentual que se consiga dizer? Ou não, é mesmo pouco e apenas um pontapé de saída?

P2: A proposta contribui bastante e é muito aproveitada para o relatório inicial. Principalmente porque, no caso dos clds, investimos bastante na proposta e foi um trabalho mto aproveitado para o relatório inicial. Mas é curioso que nos últimos cadernos de encargos que tenho visto é pedido que não apresentemos a TdM. A nossa interpretação é que, garantidamente, não é na perspetiva das equipas de avaliação terem pouco trabalho com a elaboração da proposta. Acho que não é para nos poupar... Mas tendo em conta que o processo de discussão da tdm é um processo que tem os seus gastos de tempo e de recursos, tecnicamente poderá ser. Não tenho uma perspetiva clara. Apresentar logo a perspetiva da tdm na proposta, pode ser uma forma de ver se a equipa compreendeu o objeto de avaliação. Nunca pensei mt sobre isso. Vejo as duas perspetivas: as equipas não avancem para uma coisa que pode estar ao lado; mas podia haver um ganho em avançar-se qualquer coisa. Mas reparámos já nisto, em cadernos de encargos que diziam explicitamente que não queria tdm.

P1: Outra coisa que tenho notado nos novos cadernos de encargos, é que a metodologia já vem muito mais restrita. Ou seja, basicamente no caderno de Encargos já quase dizem a quem é que querem que se faça inquéritos, entrevistas... A estrutura de proposta já está mt balizada, com tabelas que têm de ser apresentadas, matrizes completas para preencher.

Isso é vantajoso ou não, esses cadernos de encargos mais fechados?

P2: para quem vai avaliar as propostas deve ser mais fácil, pq esta tudo parametrizado. Para quem apresenta a proposta, podíamos pensar que o tempo de imaginação para responder é diminuído. Mas a criatividade perde-se, talvez também a inovação.

Os cadernos de encargos são desenhados pela AD&C, mas não são eles que gerem as medidas; eles têm um papel de gerir as avaliações, mas não gerem as medidas. Eles estão a fazer uma projecao sobre o que é pertinente avaliar nessas medidas, digo eu. Depois vocês vão ao terreno falar com que gere as medias. Não há uma descoincidência entre que o que ADC acha que é importante, com o que á aplicação das medidas no terreno? E há capacidade, ou flexibilidade, no decorrer da avaliação, de ajustar face à realidade que se encontra, aceitar descobertas que saiam fora do que está parametrizado na avaliação?

P2: acho que nunca me deparei com esse obstáculo.

P1: até mesmo ajustes na metodologia sempre se fizeram, desde que mt bem fundamentado. Por exemplo, no CLDS, pensamos fazer questionários a beneficiários de medidas mais antigas, mas houve casos em que não se conseguiu, então justifica-se.

Porque já foi há muito tempo, já se perderam os contactos, ou as pessoas estão mt dispersas no território. Há abertura, desde que justificado devidamente.

Apesar dos cadernos de encargos serem cada vez mais fechados, há possibilidade, por um lado, de a equipa justificar que não vai seguir dado caminho, mt bem justificado; e, por outro lado, sugerir algo para colmatar essa “falha”; e para acrescentar, para além de cumprir com o que é solicitado?

P2: quando a proposta é feita, é avaliada e temos acesso à fundamentação e avaliação da mesma. Por vezes, dessa avaliação, surgem sugestões relativas à metodologia. Por exemplo, em vez de entrevistas, podia ter sido proposto focus grupo, pq é mais enriquecedor. Há uma reunião inicial com quem encomenda o estudo e no relatório inicial já está esta reflexão das recomendações, desde que a equipa de valiação tb tenha concordado com. Achoq eu há flexibilidade de ambas as partes, para adaptar a metodologia, quer no relatório inicial, reflexo dessa avaliação da proposta, ou depois de análise de dados e de se ver que dada técnica não é a mais adequada.

No caso dos CLDS, achávamos que havia informação em bases de dados sobre os projetos, mas depois vimos que não: o que existe são dossiers físicos. O que está em bases de dados não permite analisar informação rica sobre os projetos em si, que acoes foram feitas, não está sistematizada, esta informação. Mesmo para ver os dossiers foi complicado, tivemos de ir para o ISS. TUDO EM PAPEL!! Incluindo relatórios de progresso, suas avaliações, relatórios intermédios, etc. tudo compilado em papel e que não foi passível de tratar. Então a análise cingiu-se a dados que estavam disponíveis e que não correspondiam à expectativa inicial.

Depois o acesso à informação que se recebe, de dados quantitativos, que envolvem bases de dados que envolvem mts atores, isto é complicado. A avaliação é de uma medida, mas como tem mts especificidades, mts bases de dados, sistematizada de formas diferentes, não é fácil compilar tudo e tratar.

Nestas situações, a entidade adjudicante compreende a situação e essa é a justificação: a inexistência de dados, que justificam não cumprir com qualquer coisa que foi em sede de proposta, ou no relatório inicial.

Como é que isto acontece? Pq que é tao complicado aceder a dados?

P1: no caso dos CLDS acho que o ISS demorou mt a responder. Mas noutras avaliações, claramente é mau funcionamento sistemas de informação, má gestão dos programas. Dificuldade para extrair dados, que saem com erros. As vezes estamos a trabalhar dados e depois veem que estão mal, enviam novamente, depois faltam variáveis. Acho que os sistemas de informação são mt difieis.

P2: a questão da proteção de dados tb agora tornou td mais difícil.

P1: é preciso tornar os dados anónimos, mas acho que é mt problema do pp sistema.

P2: neste caso era ADC, e acho que eles até vão articulando atempadamente com os POs, já dizendo que a equipa de avaliação vai arrancar na data x, que vao precisar de dados na data y, mas mesmo assim é difícil. Às vezes levamos meses à espera de informação e a receber aos poucos. Depois, às vezes, temos de ir para instalações próprias para recolha de dados...

Ainda sobre a TdM: a tdm que é construído envolve o desenho do programa, mas depois na avaliação, nos resultados da avaliação, o desenho do programa é objeto de avaliação, ou centra-se mais nos mecanismos de mudança, nas realizações, ou seja, há uma avaliação do desenho do programa? Avalia-se a relevância e a coerência do programa?

P2: tudo o que é as dimensões ou critérios de avaliação, como eficiência, eficácia, é direcionado para o programa, é isso que é avaliado. No caso dos CLDS, foram identificadas áreas que devia ter havido intervenção e não houve, avaliou.se a coerência e a relevância. A partir de dada altura a TdM e a articulação com a entidade adjudicante é mt mais centrada na parte do programa.

Como encaram os resultados das avaliações: qd chegam ao fim, chegam a resultados ou não? É que parece haver uma certa insatisfação da comissão europeia com as avaliações feitas em PT...

P2: em relação aos CLDS... há avaliações com temáticas mais queridas que outras. O processo e o resultado da avaliação dos CLDS, pessoalmente, deixou-me bastante satisfeita. Acho que há um problema com os timings das avaliações. Parece que estão a ser feitas em alturas em que os resultados delas depois tb não são propriamente bem canalizados. Começas a fazer a avaliação qd já tens resultados que permitam construir a nova geração, qd ja há um novo quadro. Isto tb poe pressão para ter resultados mais rápidos.

Alguma questão de avaliação mais difícil de responder?

As duas: talvez o VAE, é sempre um mistério.

Há um problema metodológico ou operacional nas avaliações, no sentido de não se encontrar o melhor caminho para fazer ou se é um problema de interpretação...

P1: acho que em relação à TdM, desde os CLDS, que há a questão de diferentes interpretações do que é a TdM e que elementos devem estar na TdM. Isso leva a que as TdM e os relatórios iniciais andem sempre para trás e para a frente até se chegar a um consenso entre as diferentes conceções do que a TdM e das medidas as que devem estar na TdM.

Sobre a gestão dos projetos e o triangulo que referiram. Qual é a dimensão que tenta mais compensar as outras? É certo achar que é o custo é a dimensão mais reduzida e as outras é que compensam?

P1: sim, geralmente o custo nunca se altera. O orçamento não se reajusta. Geralmente o que fazemos são ajustes no tempo, ou nos RH, que tem uma implicação no custo para a entidade, mas não na avaliação. O tempo é que sofre os ajustamos, pq tb não há mts alterações no âmbito. O âmbito e o custo para a entidade adjudicante não mudam. O tempo é o que mexemos mais.

Já se falámos mt sobre TdM, mas ainda uma questão: qual o papel da TdM nas avaliações de políticas públicas?

P2: a TdM supostamente é o guião da avaliação, é o caminho que tem de ser percorrido e que serve para validar se foi ou não percorrido.

P1: mas apesar disso, a TdM não esgota toda a avaliação. As vezes o contributo da TdM para aquilo que são os objetivos da avaliação até nem é mt. Acho que se perde mt tempo a definir a TdM e o o seu contributo para os resultados da avaliação acaba por não ser tao abrangente como isso. Por exemplo, nós temos uma avaliação, pedem TdM e um conjunto de QA com subquestoes. Fazemos a TdM e ela dá para responder a apenas 1 ou duas QA; depois há todas as outras QA que a TdM não dá contributos. Há um grande investimento. Não sei se a TDM é sempre a metodologia mais adequada para aquele tipo de avaliação. Por exemplo, nos relatórios iniciais há mta discussão em torno da Tdm, as caixinhas, os mecanismos, mas no relatório final a Tdm fica restrita a poucas QA qd há mts. A TDM não responde a tudo.

P2: responde a questões de coerência, de eficácia, de impacto, eventualmente.

P1: sustentabilidade às vezes tb não dá muito. Eficiência não dá nada.

P2: na eficiência pode dar ou não... depende. E depois temos o VAE.

P1: esta é a sensação com que fico: faz-se a TdM, constroem-se os instrumentos para a TdM e depois há QA que são respondidas por coisas que não estão diretamente ligadas com a TdM.

Como superar isto?

P1: dar mais tempo para o relatório inicial, para se desenvolver bem metodologicamente. As TdM virem já predefinidas no Caderno de Encargos tb vai acelerar um pouco o processo. Pq ADC eja tem ideia do que quer e assim não se perde tempo em adaptar a nossa visão à deles. Já haver um pre-esboço, então a equipa já tem ideia daquilo que se espera. Na verdade, a TdM deve ser feita pelo pp programa. Mas os que não têm, haver uma primeira definição é um bom ponto de partida para as equipas. Isto faz perder menos tempo em conciliar as visões.

Se fosse hoje, e pensando no caso dos CLDS, o que fariam de forma diferente? Ao longo das avaliações que têm vindo a fazer, que aprendizagens, mudanças, que balanço?

P2: de proposta para proposta procurámos ao máximo desenvolver a TdM na proposta. Acho que foi uma estratégia às vezes ingrata, pq se não ganhamos a proposta já investimos imenso, mas tentamos que com isso ganhar a proposta e estar o mais próximo possível do produto, do relatório inicial. Tem o outro lado, se não ganharmos a proposta já investimos mt. Mas isto foi algo que fomos desenvolvendo.

Uma aprendizagem, sempre que possível, é ter nas equipas as pessoas comprometidas com a avaliação, com a proposta, e serem as que vão ficar a trabalhar no terreno, que têm o know-how. O resto não depende de nós.

O prolongamento para o relatório inicial é importante, nem sempre suficiente. Ele devia chamar-se relatório essencial! Tem uma importância tal que não se comporta num mês, está muito descompensado. Um projeto de 12 meses desenhado num mês, 30 dias em 365?? Para apresentar a teoria, os instrumentos adequados à teoria? E o hiato temporal entre concorrer-se, adjudicar a arrancar, pode perder-se as pessoas que estavam na proposta. E, entretanto, é preciso reunir com todos, todos têm de estar orientados para o mesmo, para se começar a trabalhar. É preciso mais tempo para o relatório inicial, é isso.

Como é que é essa negociação dos tempos?

P1: falando nos CLDS, e em termos gerais, sempre houve abertura para estas negociações. Até pq mtas vezes as demoras não são só do nosso lado. Por exemplo, entregamos a primeira versão do relatório inicial e depois esperamos mt tempo para receber feedback, depois fazemos alterações, temos de esperar novamente feedback... às vezes são mt pormenorizados nas análises que fazem e tb nao tb levam mto tempo. Então nunca foi difícil negociar os prazos. Ultimamente tem havido maior insistência para os prazos, nas últimas avaliações. Nas primeiras houve avaliações que descarrilaram mt no tempo. Também pode ter que ver com fechar o quadro comunitário.

P2: tentamos sempre ter uma boa relação. Mas diria que às vezes era desnecessário o sentimento de incumprimento nas equipas, pq é injusto, de algum modo. Sabemos que não vamos cumprir aquele prazo, pq não é exequível e sabemos à partida que não vamos cumprir.

P1: Nos CLDS, estivemos mt tempo à espera de bases de dados. Houve um problema de comunicação, pq achávamos que havia as bd com essa informação, mas não havia e não nos diziam claramente, então estávamos sempre a pedir. Só dps de irmos ao ISS e vermos é que percebemos, pq nem o PO, nem o ISS nos dizia que não tinha essa informação. Isto prolongou tudo mais no tempo, para a avaliação.

P2: portanto, o valor atribuído às avaliações melhorou nos últimos anos, mas na altura dos CLDS, era menor e insuficiente para o que era pedido.

Relatório/ transcrição das entrevistas a responsáveis nas estruturas dos programas em avaliação

Gostávamos de saber um pouco sobre o POCH e a primeira questão que gostávamos de colocar é se houve um diagnóstico prévio à criação e implementação do programa?

P1: Era uma obrigação regulamentar, do programa, fazer uma avaliação ex-ante, para a sua aprovação – isto Naquela época, porque atualmente já não é uma obrigação, pode se fazer, mas já não é uma obrigação. Essa avaliação é ex-ante fazia parte do próprio processo de programação. Portanto aquilo que foi o racional do programa passou por esse escrutínio; a avaliação fez um conjunto de recomendações, algumas foram seguidas outras nem tanto, o que teve também a ver com algumas das opções políticas, porque no final do dia, a opção é sempre política e não meramente técnica.

P1: Há uma nota que vos queria deixar, (...), não é uma avaliação do POCH, e esse foi um dos grandes desafios que nós tivemos na programação, porque todas as avaliações que nós fizemos foram de caráter temático, há uma única avaliação que está agora a ser desenvolvidas que é do POCH, e é uma avaliação da obrigação regulamentar, E obriga todos os programas a ter uma avaliação conjunta de todo o programa e que vai beber muito às avaliações temáticas, mas o que nós fomos promotores e coordenadores foram avaliações de caráter temático, uma das quais objeto desta entrevista. a avaliação que aqui foi feita tinha a ver com uma das grandes frentes do programa, na ótica da formação inicial de jovens, no combate ao insucesso escolar, ao abandono, a promoção da empregabilidade, por isso também incluiu a avaliação do emprego jovem e dos jovens NEET. E foi esse o objeto de avaliação que foi avaliado, mas que abrangeu não só o programa operacional de capital humano, como também outras intervenções, sabendo claro, que o programa tem uma intervenção muito predominante sobretudo nas regiões que financia tipicamente (Norte, Centro e Alentejo), Mas mesmo nessas regiões as intervenções foram suplementadas por intervenções dos programas operacionais regionais (...). [Portanto, a avaliação avaliou componentes do POCH e dos programas operacionais regionais]. Isto implica, do ponto de vista da arquitetura de todo o modelo da avaliação, que seja um modelo muito mais exigente. Exige envolver todas as entidades de gestão, desde o início do processo de programação e planeamento da mesma. O que passa para o exterior é que isto é uma avaliação do POCH, porque é quem coordena a avaliação, e é o maior investidor que está a ser avaliado do ponto de vista dos fundos europeus, mas estamos a avaliar o conjunto das políticas de Portugal 2020, e incluiu também as regiões autónomas. No nosso plano de avaliação tentámos focar sobretudo na coordenação dessas avaliações de caráter temático (...).

P1: Todos estes processos de avaliação tiveram como objetivo e preocupação o contrafactual. As avaliações baseadas na teoria, com o contrafactual, em todas elas foi feito uma avaliação de impacto, ou seja, à data, quando se concebeu o diagnóstico e a perceção que nós tínhamos do nosso lado era de que o grande valor acrescentado - sobretudo Hoje, em que os sistemas de informação estão cada vez mais desenvolvidos, como a consistência que não tinham há uns anos atrás, em que a própria regulação comunitária nos obriga desde logo ao escrutínio do ponto de vista da monitorização, com os indicadores comuns comunitários e todas as exigências nesta matéria; ou seja nós já tínhamos um conjunto de informação, de reporte periódico, os boletins que vamos fazendo, que nos obrigam a uma lógica de monitorização e assim irmos acompanhando e fazendo alguns acertos na própria intervenção dos programas, sempre que as coisas possam estar a não correr tão bem, como era esperado quando as coisas foram planeadas. (...). Aquilo que se entendeu que era o grande valor acrescentado que estas avaliações nos podiam trazer tinha a ver com os impactos, estimar qual era o contributo dos programas. (...). E o contratual entra aqui, ser mais uma metodologia, um complemento às metodologias baseadas na teoria para procurar identificar mais possível ou procurar comparações... por exemplo, quais as diferenças de desempenho entre quem frequenta uma medida financiada e quem não frequenta essa medida, ou que está noutras medidas. ISTO é uma dificuldade especialmente na área da formação inicial de jovens, porque não temos quem não tenho sido abrangido por algum tipo de medida, a questão é essa. O que fizemos foi comparar perfis comparáveis, por exemplo as vias mais profissionalizantes com as vias mais científico humanísticas e de ver qual foi o desempenho e que inserção no mercado de trabalho. (13'30)

Voltando um pouco atrás, naquilo que nos contava sobre a avaliação ex-ante e de que forma é que ela depois marca a construção do programa e este trade off entre propostas talvez técnicas e propostas políticas, este confronto entre os resultados do trabalho técnico, onde o contexto político marca com certeza, e muito.... Pode-nos falar um pouco sobre como é que isto impactou no desenho do programa, de forma mais concreta?

P1: Tenho alguma dificuldade em falar de, porque nessa fase não estive tão envolvido. Os processos foram basicamente em paralelo, ou seja, a avaliação ex-ante correu ao mesmo tempo que estava a fazer a programação, e houve algumas dimensões e recomendações que a avaliação referiu, pelo que vi do programa, penso que foram ponderadas na programação. (...).

P1: Houve preocupações que avaliação também suscitou e que foram, de alguma forma, ponderadas até num conjunto de apoios que depois foram sendo colocados no terreno. (...).

E no desenho, depois, do próprio programa, estava previsto um sistema de monitorização, de implementação? Como é que têm decorrido os vários eixos, neste sentido?

P1: Nos programas há questões transversais, desde logo a aplicações regulamentares, que são muito exigentes, até pelo nível de escrutínio que têm tido ao longo deste tipo de programação. eles estão associados, alguns desses níveis de utilização, a uma coisa chamada quadro de desempenho, que ocorreu em 2018, ao qual está associado um conjunto de compromissos. Neste caso, até eram fundamentalmente de realização física e financeira. Portanto, desde o início que a questão da monitorização era uma questão específica, e fundamental por causa desta dimensão, do quadro de desempenho. Tivemos muito trabalho em 2018 e vamos ter agora, com o encerramento do programa em 2023 a 2025. Quanto a esse nível de desempenho é preciso ter algum cuidado, porque pode ter consequências financeiras.

Permita-me só interromper só para ter a certeza que compreendo. Talvez a minha pergunta também acabou por ser um pouco ambígua neste aspeto: esta monitorização que estamos a falar e este quadro de desempenho é para o próprio POCH, certo?

P1: Sim.

E no que respeita às entidades beneficiárias: nas medidas que desenvolvem, se também elas estão sujeitas a um sistema de monitorização, ou se são vocês, a estrutura do POCH que faz essa monitorização?

P1: Nós temos vários níveis de acesso ou de concretização dos indicadores, que fomos estabelecendo para o programa, como se fossem várias camadas de indicadores. Temos uma camada, que são os indicadores comuns comunitários, que são aqueles que a Comissão nos obriga a reportar, no sentido de reportar anualmente, no nosso relatório anual. Existem os indicadores que nós estabelecemos para o nosso programa, que são os chamados indicadores específicos do programa, nomeadamente aqueles que contribuem para o quadro de desempenho, tanto físicos como financeiros. E depois temos uma camada mais ao nível das entidades beneficiárias, que são aqueles indicadores que nós contratualizamos diretamente com estas entidades beneficiárias e que, no fundo, nos dizem que elas se comprometem a concretizar uma série de patamares mínimos de desempenho, que fazem com que e aquele financiamento que receberam concretize um mínimo que aquelas se dispuseram a concretizar, quando fazem este contrato connosco, o contrato de financiamento com o POCH. Esse nível de monitorização ao nível das entidades também é feito por essa via do estabelecimento de contratualização de indicadores mínimos de desempenho face àquilo que estão a receber.

E só para perceber melhor: é um mecanismo no qual, caso não haja o cumprimento, gera impactos também depois no financiamento?

P1: Sim. Sempre que não se concretize o patamar mínimo, a entidade pode vir a sofrer correções financeiras, relativamente àquelas operações, ou futuras; já depende um bocadinho daquilo que acontece, se há cumprimento ou não. Portanto, a monitorização também é feita a este nível, não só ao nível mais macro dos indicadores físicos e financeiros ao longo do programa, mas também ao nível de cada entidade que recebe o apoio.

E de que forma é que consideram, da vossa experiência, que que esses sistemas de monitorização contribuem ou têm contribuído para a concretização do próprio programa?

P1: Obviamente que é muito, é muito importante, nós poderemos fazer esta análise que é mais ou menos obrigatória, mas nós, para além dessa, fazemos nós próprios um acompanhamento mais ou menos regular, nós temos boletins publicados trimestralmente, em que fazemos o reporte da monitorização dos nossos indicadores, mas também de outros aspetos, um bocadinho mais humanizados da nossa intervenção, mas que nos dão uma sensibilidade muito grande em relação àquilo que tem sido o nosso acompanhamento.

P1: E depois há aqui outro fator. Estes dados que nós retiramos da monitorização acabam por também alimentar as avaliações que vamos produzindo ao longo do programa. É outra variável que nós conseguimos retirar da questão da motorização, porque a avaliação também se alimenta muito dos dados da monitorização. É objetivo também da própria monitorização alimentar a componente da avaliação.

E verificaram isso na avaliação que nos traz aqui?

P1: Sim. No fundo, isto é uma lógica de uma cadeia, ou seja, vou dar um exemplo muito concreto. Tipicamente nós temos 2 indicadores-resultados associados às formações que iniciamos. Um é taxas de conclusão, ou seja, quanto é que iniciaram? Quantos é que chegam ao fim? E quantos concluem dentro do período expectável para essa formação? Nós tipicamente financiamos formações longas, seja de jovens, seja de adultos. Outros programas depois financiaram formações mais curtas e que também foram objeto de avaliações. E depois a questão da empregabilidade, ou prosseguimento de estudos, que essa é outra métrica que utilizamos. Isto significou depois, na lógica da cadeia, que nós contratualizamos com cada um dos beneficiários esse mesmo tipo de indicadores em função do seu ponto de partida. Ou seja, se eu tenho como indicador-resultado do programa que quero chegar aos cursos profissionais, que tenho uma meta definida na programação, que é partir de uma taxa de conclusão de cerca de 60%, para chegar a 70% em 2023. O que se fez foi contratualizar esse objetivo com os beneficiários, mas não de forma cega, em função do ponto de partida de cada um. (...). E então, se não chegassem a esse patamar, então podiam ter algum nível de penalização, mesmo assim, limitado; inicialmente era limitado a 10% do montante e depois até foi reduzido para 5%. E em casos muito graves, ou seja, de completo desvio, ou seja, que não cumprem menos 50% que se comprometeram no limite, aí pode, pode ter consequências mais gravosas. Pode implicar até a revogação da operação, se não houver uma razão muito fundamentada que justifique o sucedido.

P1: Este racional não foi fácil, porque implicou alguma rutura, mas também uma evolução significativa face à tradição que havia na gestão dos Fundos, que foi passar isso para toda a cadeia de gestão dos fundos, desde o beneficiário até aos nossos gestores de projetos. Portanto, hoje o nosso gestor de projetos quando avalia um pedido de reembolso, uma das coisas que faz é avalia e vê qual é o ponto de situação em que a entidade está. (...). Durante a implementação dos projetos, a única coisa que as equipas conseguem sinalizar é, basicamente, é quantos alunos se comprometeram a abranger e quantos estão a abranger, portanto, a distância a essa meta. (...).

De uma forma geral, aquilo que também gostávamos de perceber é o seguinte: para esta avaliação, o caderno de encargos foi publicado pela Secretaria-Geral de educação em ciência. Nós gostávamos de perceber qual foi o papel da estrutura do POCH no desenho desse caderno de encargos e a própria expectativa que tinham em relação à avaliação. Portanto, se tiveram um papel no desenho do caderno de encargos, e de que forma é que a esse papel se se desenrola, que expectativas, que contributo para o desenho da metodologia, para os objetivos...

P1: No caderno encargos só aparece a Secretaria-Geral por uma razão muito simples. Do ponto de vista operacional, porque as autoridades de gestão são estruturas de missão, não têm autonomia nem financeira, nem administrativa, nem coisa nenhuma. E, portanto, as especificações técnicas, na parte jurídica, a Secretaria-Geral tem uma experiência de contratação pública muito grande, pelas suas próprias funções, que têm de dentro da orgânica do Ministério da Educação e na área da ciência.

P1: E, portanto, a componente das especificações técnicas e dos cadernos encargos, nomeadamente da componente mais técnica do que é que era o exercício de avaliação é da nossa responsabilidade. Quando digo nossa, é ponto de vista do processo de coordenação, desde logo coordenação, porque há já, no contexto da própria rede para a monitorização e avaliação que devem conhecer, portanto, há um conjunto de orientações, documentos gerais de orientação para a avaliação que são discutidos nessa rede, onde nós também participamos, que é uma rede coordenada por ADC. E, portanto, há um conjunto de

orientações associadas ao próprio ao próprio processo do Plano de Avaliação, portanto, há algumas das questões gerais que já são utilizadas, e que são transportadas para os cadernos encargos e, em particular, para as especificações técnicas dos estudos, porque são matérias em que se procura uma coordenação global no contexto dos fundos, ou seja, quer do ponto de vista da estrutura, da existência de grupos de acompanhamento, ou seja, um conjunto de matérias que são transversais a todos os exercícios de avaliação dentro do plano de avaliação dos programas, que por sua vez está integrado num plano de avaliação mais global, que é o plano global de avaliação do Portugal 2020. E que tem um conjunto de orientações gerais e são discutidas nesta rede e, portanto, isso também se traduz nos cadernos de encargos, sem prejuízo de especificidades próprias de cada objeto. Mas há um conjunto de matérias comuns, em matéria de comunicação dos resultados da avaliação, ou seja, há um conjunto de matérias que são relativamente coordenadas e articuladas nessa sede e com um documento que existe, que é um guia, no fundo, que as autoridades de gestão depois utilizam para lançar os seus próprios processos de avaliação, que tem um conjunto de orientações gerais e depois são traduzidas nesse.

P1: Mas o caderno de encargos na prática é da nossa responsabilidade e também procurando logo envolver na fase de conceção desse caderno de encargos outras autoridades de gestão e outros organismos que estão ligados àquela política pública, ou seja, que já são envolvidos na própria preparação dessas especificações técnicas e depois são entidades que integram os grupos de acompanhamento das avaliações. 32'40. (...).

Uma vez que o caderno de encargos é uma responsabilidade do vossa, qual é que consideram que é o papel da teoria da mudança numa avaliação ex post?

P1: A teoria da mudança no fundo é uma metodologia que nos obriga a olhar para um objetivo e depois perceber que mecanismos é que nos levaram a esse objetivo, se esse objetivo se concretizou, e se esses mecanismos nos levaram a essa concretização. Do meu ponto de vista, e isto é um bocadinho transversal às avaliações de impacto, é que faz sentido olhar para as avaliações com esta lógica, de perceber onde é que chegámos com a intervenção que foi feita. Outro, e se calhar vamos falar sobre isso, é o aspeto da complementaridade de outras metodologias. E acho que aí é que entra a questão do contrafactual, porque nos permite ao mesmo tempo perceber as alternativas que poderia haver a esta teoria, ou pelo menos às teorias que estamos a aplicar à política pública. A complementaridade entre as 2 metodologias, acho que é muito rica ou enriquece muito os processos avaliativos.

Aquela avaliação que houve, ex-ante, já tinha também uma teoria da mudança desenhada?

P2: Eu não sei, não acompanhei, não estava ainda cá. Eu diria que sim, mas não tenho presente, tinha que confirmar.

P1: Do que me recordo, sim, embora não tão forma tão explícita como, entretanto, ganhou. Mas prática acabou por essa componente.

Qual é o papel da teoria da mudança numa avaliação ex-post? (37'00)

P1: essa é uma pergunta difícil. A teoria da mudança tem sempre relevância. Toda a lógica de intervenção de um programa, uma intervenção pública tem uma lógica, algum tipo de objetivo de mudança, se não, não vale a pena gastar dinheiro público, não é? Neste caso, portanto, espera-se que o investimento que é feito produza algum efeito, alguma mudança estrutural. Aqui, trata-se de Fundos que têm como grande objetivo estratégico fazer alguma alteração estrutural naquilo que é a situação do ponto de partida em relação ao ponto de chegada. E, portanto, sabendo que depois, ao longo do tempo, isso pode ter ajustamentos, até porque as circunstâncias mudam, como foi agora com a crise pandémica, com aquilo que foram os impactos e outros que ainda não estamos a sentir daquilo que é esta guerra na Ucrânia. Mas o exercício de programação parte do pressuposto que nós vamos alocar recursos públicos para mudar alguma coisa. Face ao contexto, o exercício tem de ser revisitado e às vezes obriga a calibrar aquilo que é a programação inicial para responder a necessidades mais conjunturais, face a crises económicas, como na crise de 2008, na crise pandémica e agora com a crise devido a guerra na Ucrânia. Mas fazê-lo sem perder o foco estrutural! Ter de conciliar os dois objetivos (conjuntural e estrutural) nem sempre é fácil. E desse ponto de vista, na avaliação expost, é sempre ter presente o que é a programação, para que é que os recursos são, pelo menos inicialmente, planeados, quais foram os ajustes que foram necessários fazer a esse planeamento inicial, por via da mudança

das circunstâncias, e em que medida é que depois essas alterações que as circunstâncias exigiram usaram ou não, ou em que medida levaram a um equilíbrio. (...). Por exemplo, na pandemia, os Fundos foram usados para manter os empregos das pessoas, quando tipicamente os fundos estão concessionados para promoverem a criação de emprego e emprego mais qualificado. Mas a pandemia, por exemplo, obrigou em boa medida, a alocar uma grande quantidade de recursos públicos, inclusive os fundos europeus, para aguentar o emprego, para que as pessoas não fossem despedidas.

A nossa questão com o papel da teoria da mudança neste contexto de uma avaliação ex-post é precisamente o duplo papel do avaliador, da equipa de avaliação desenhar e testar a teoria da mudança, não é? E perguntamos isto e perguntávamos também se tinha havido uma teoria da mudança na avaliação, na medida em que a teoria da mudança ser desenhada e depois testada pela mesma equipa pode gerar algum enviesamento dos resultados ou será uma das eventuais limitações dessa situação, de ter esse duplo papel. E, portanto, como é que vem esta possibilidade? Como é que, naquilo que foi o exercício desta avaliação, como é que essa articulação foi feita, entre o desenho e o teste, sendo que é mesmo a equipa que o está a fazer?

P1: A única resposta que eu tenho para isso é que o exercício de avaliação tem por base metodologias, que têm métodos científicos por trás. Eu vejo as avaliações como exercícios de investigação, embora, embora muito operacional. Têm metodologias, têm métodos científicos e, portanto, procuram ser o mais objetivas a processos de intuição, com processos de entrevista, portanto, têm metodologias próprias e que procuram a partir dessas metodologias produzir factos. Aliás, é uma das nossas “Guerras” entre aspas, no bom sentido da palavra, uma questão sempre presente em discussão com as equipas de avaliação, é exatamente de que aquilo que são as conclusões das evidências estejam devidamente sustentadas nos factos que correm do processo de pesquisa que a avaliação traz, com base nos métodos de avaliação a que recorrem. (...) Nós não queremos avaliações “achómetro”, nem mediáticas, nós precisamos de exercícios de avaliação para termos evidências devidamente fundamentadas e sustentadas em métodos de avaliação, que métodos de investigação científica associada às ciências sociais, à economia, (...), portanto, sustentadas em evidências que resultaram de um processo de investigação. E, portanto, independentemente de, no limite, esse processo ser feito por que esteve na avaliação ex-ante – embora tenhamos algumas precauções, porque às vezes os avaliadores estiveram no processo de programação, e não queremos que depois apareçam no lado da avaliação. No limite, a equipa que fez a ex-ante pode estar na ex-post, ou ongoing, porque se baseia em métodos e metodologias de investigação e, portanto, há um rigor e independência que é assegurada por essa via, acho que temos segurança de isso não cria esses enviesamentos. (...).

E no caso da teoria da mudança, e em relação a esta avaliação que foi feita, que critérios de avaliação era expectável que a teoria da mudança desse resposta?

P1: No nosso caso, claramente, estamos a falar de avaliações de impacto e responde basicamente a 2: eficácia, qual é o nível de eficácia que as medidas que foram financiadas tiveram naquilo que eram os objetivos fundamentais da política naquelas áreas: produção do sucesso, combate ao abandono, e maior empregabilidade dos jovens. E uma dimensão tipicamente sempre mais difícil de se ter resultados mais objetivos, a componente da eficiência. Eles executados, mas podia ou não ter atingido os resultados com mais ou menos recursos do que aqueles que utilizei? É onde temos mais dificuldade em termos evidências mais substantivas na resposta a essa questão. Há uma terceira dimensão, mas é uma coisa muito específica, porque estamos a falar da avaliação dos fundos comunitários, que é o valor acrescentado dos Fundos: O que é que os fundos trouxeram de valor acrescentado? O que é que eles permitiram fazer que se não houvesse fundos, não ia acontecer ou depois não ia acontecer com a dimensão que foi possível (...). É uma resposta fundamental para evitar ou atenuar aquilo que corre muito na opinião pública, e até política, de que os Fundos funcionam para substituir meramente o orçamento de Estado, não é? 49’00. (...)

Neste processo da construção da teoria da mudança, e na relação com a equipa de avaliação enquanto estrutura responsável pelo programa, quais foram na vossa perspetiva os principais desafios encontrados nesta avaliação, para o desenho da teoria da mudança?

P1: Principais desafios... Há aqui uma questão que é preciso chamar a atenção: os próprios exercícios de programação também têm vindo a ser aprofundados e melhorados, já exigem de alguma forma que essa teoria da mudança esteja, de forma mais

explícita ou menos explícita, nos próprios exercícios de programação. Vou dar um exemplo, a programação do Portugal 2020; partia de um racional em que se questionava: quais são os problemas fundamentais? Quais são as medidas que eu vou tomar e com isso, no final, o que é que eu vou procurar mudar? Portanto, com isso, define, qual é o problema fundamental, os objetivos específicos, o que é que eu vou querer mudar durante os 7, 8, 9, 10 anos de intervenção, (...) e no final dos resultados é que eu quero atingir. Portanto, isso já está um bocadinho no ADN da própria programação. Com mais ou menos sofisticação, com o melhor ou pior capacidade, são as famosas cadeias de programação e, portanto, é sobre essas cadeias de programação, desde logo se pede que as avaliações ex-ante se pronunciem: a programação está a ser bem conseguida, não está a deixar coisas de fora, está a deixar coisas de fora com aquilo que está a ser programado para responder ao problema, é suficiente ou dá garantias de ser suficiente, que permitam atingir aqueles resultados? No fundo, é este questionar do próprio início de programação que está desde logo no exercício da avaliação ex-ante e que vai acompanhando a própria implementação dos programas ao longo. Nós estamos agora em pleno início da programação do PT2030 também há essa lógica, de alguma forma está a ser seguida. (...).

P1: Sabemos que também depende muito mais do grau de consistência, de capacidade da própria administração pública para definir essas políticas públicas que são objeto de financiamento e que estão a ser programadas, sabendo são sempre a médio-longo prazo.

P2: já agora, para acrescentar, porque isso também é relevante para aquilo que era a pergunta anterior, sobre o enviesamento entre o desenho da teoria da mudança na ex-ante e na ex-post; porque, na verdade, a política pública, seja ela qual for, apesar de muitas vezes não ser explícita, tem subjacente uma teoria da mudança quando se quando ela é desenhada e, portanto, o que se quer da avaliação é perceber se essa teoria da mudança que era desenhada para a política pública corresponde ou não àquilo que se tinha pensado quando o legislador pensou resolver um problema e quis formular uma política para resolver. E o que o que P1 está a dizer é importante. Nós, quando agora estamos a desenhar o próximo período programático, já temos em consideração uma série de objetivos que queremos cumprir, e isso faz parte de uma teoria da mudança para as políticas que queremos implementar, e isso tem contributo para aquilo que vai ser a teoria da mudança da avaliação dessas mesmas políticas, durante o período em que ela está a ser implementada e depois mais à frente, perceber quais é que foram os efeitos da própria implementação da política. A avaliação, na verdade, deve responder, no meu ponto de vista, também deve responder àquilo que é o desenho da própria política; neste caso uma coisa mais estruturante, sempre com a lógica incremental que falávamos há bocadinho, desde as mudanças, ao que é que é de conjuntura, o que é que é estrutural, choques externos, seja de uma pandemia, seja de uma guerra, seja do que for, é preciso sempre fazer este acompanhamento da política. Mas, no fim, perceber que esta correspondência entre estas teorias da mudança, por assim dizer, é muito relevante. E isto acaba por por atenuar esse eventual enviesamento, que poderia eventualmente existir, seja por uma equipa de avaliação seja por outra, mas como estava a dizer, nós somos um país relativamente pequeno, as pessoas acabam por estar todas muito próximas; os peritos, os especialistas acabam por ser todos muitas vezes os mesmos e é difícil as pessoas não estarem todas envolvidas, seja no período do desenho, seja no período da avaliação, ou seja, em ambos é muito difícil não serem, porventura, pelo menos parte dessas pessoas serem as mesmas.

O que nos contam também é que, no fundo, ao longo dos últimos anos essa teoria da mudança tendencialmente, ela está cada vez mais explícita, não é? Quer dizer, porque sabemos, claro, há sempre um racional...

P1: Claro, há sempre uma teoria de mudança. Ela não está expressa na legislação, mas há sempre uma teoria da mudança que subjaz a qualquer política pública: onde é que nós queremos chegar com esta intervenção? Isto é uma é uma teoria da mudança. Que que mecanismos vamos mobilizar? o que é que vamos fazer? como vamos partir esta esta intervenção em várias componentes para atingir este fim? Que recursos vamos mobilizar? Como é que vamos monitorizar ao longo do tempo? Tudo isto faz parte de uma teoria da mudança, ela não é expressa tal e qual como nós conhecemos a teoria da mudança do ponto de vista teórico, mas o que está subjacente é exatamente isso. Eu até diria que todas as políticas públicas deveriam ter expresso uma teoria da mudança que fosse evidente para qualquer pessoa.

No fundo é isso, porque ao ser a equipa que faz esse apanhado, que faz esse desenho e que depois vai testá-lo, há aqui sempre a possibilidade do desenho ser sempre muito reajustado àquilo a que se se chega...

P1: Parece-me que até faz sentido que assim seja, é preciso que seja claro, que dê resposta, que encontre esses mecanismos de mudança. Uma avaliação ongoing, ou a própria monitorização, também tem esse papel, diria, de olhar para essa necessidade ou não de ajuste, de retrocesso, de ajustamento numa outra direção, acho que faz parte do próprio ciclo da política. E tendencialmente apoiamos muitas medidas políticas que são estruturantes, que se concretizam ao longo de vários de vários anos e tendencialmente com continuidade, portanto, mais necessidade de ajustamento ao longo do tempo. Sofremos choques externos e a própria sociedade também evolui, portanto, é preciso irmos fazer ajustamentos. (...) Do ponto de vista da programação, já vamos olhando para o processo de uma perspectiva muito mais precoce do desenho da teoria da mudança, que depois seja mais fácil, ou mais à frente, concretizar uma teoria da mudança na avaliação. 57'00. (...).

E no caso, depois, da avaliação de impacto contrafactual, a que critérios de avaliação é que era expectável que esta avaliação de impacto respondesse?

P1: Eu penso que há aqueles que há pouco tinha enunciado, ou seja, é basicamente a questão da eficácia (...). Aqui a dificuldade é identificar os que foram e os que não foram apoiados, pq a maioria teve apoio publico. O que se fez foi comparar quem teve determinado tipo de tratamento com teve outro tipo de tratamento, neste caso de formação, e ver o resultado dessas formações. (...). [Queríamos], com o contrafactual, ter uma evidência mais sustentada de qual é que é o diferencial do facto de apostar numa via ou noutra, o que é que isso produz de diferencial nos percursos de vida das pessoas. Trabalhamos com financiamento que primeiro tem impacto sobre pessoas e que depois também tem a níveis territoriais e outros mais alargados, mas em primeira linha intervenção sobre pessoas. E, portanto, o que é que isso produziu depois a seguir, em matéria de desempenho dessas pessoas, do ponto de vista pessoal e também participação social, por via do aumento das qualificações.

Pensando ainda nos critérios de eficácia, eficiência, valor acrescentado europeu, sustentabilidade... de alguma forma este exercício de avaliação de impacto contrafactual contribui para algum outro critério?

P1: Eu diria que é basicamente para esse; aumento da sustentabilidade também, porque também com isto nós podemos fazer exercícios de avaliação mais ao longo do tempo (...). Também tem que ver com o próprio contexto, a própria lógica de gestão dos fundos, que tem estas lógicas de programação. As nossas avaliações são muito associadas aos ciclos de programação e, portanto, agora pedem-nos para avaliar este, a seguir vai fazer-se fazer uma avaliação ex post do outro, e a seguir já se tem de planear a avaliação do próximo, portanto, são muito marcados por ciclos de programação. Mas são muito importantes, porque depois no final do dia, há diferentes perceções dentro do próprio universo às vezes sobre a pertinência ou a validade desses exercícios para o próprio processo de programação... estamos a procurar demonstrar isso, nesta programação, que é, nós estamos a procurar encontrar a programação que vamos fazer agora, exatamente nalguns resultados não só da educação, mas das avaliações que fizemos, portanto, porque é isso que nos ajuda a sustentar algumas opções de programação e de investimento, e de priorização dos recursos públicos, porque estamos a meter o dinheiro em sítios onde temos evidências de que ter o dinheiro ali produz resultados. E os resultados são positivos dentro daquilo que é a tal mudança estrutural que se quer para o país. E acho que as avaliações nos dão essa consistência adicional, para além dos dados de monitorização, de permitir evidenciar que o país está agora a programar para os próximos 7 anos e que está a programar em função daquilo que são, não apenas decisões políticas, mas que tem também por base uma componente técnica de evidência: estamos a pôr os ovos nos sítios em nos mostram que têm alguma capacidade.

De que forma é que realmente aqui a estrutura do POCH articula com as entidades que depois executam as medidas, que são financiadas, e tendo em conta as conclusões e as recomendações das avaliações? De que forma é que depois essas conclusões e as recomendações das avaliações impactam em quem executa, em quem é financiado e executa?

P1: Nós temos neste momento definido, até devido ao processo em rede, uma metodologia e um processo depois de follow-up das avaliações, ou seja, para garantir que as avaliações não são colocadas na gaveta, que se discute as conclusões e explicações no momento em que se discute, e depois, a partir daí, mete-se coisa na gaveta e acabou. Portanto, há um processo hoje instituído, que tem tido as suas dificuldades de entrar no terreno, junto das entidades que refere, ou seja, as entidades que gerem a política pública – e também não escondo, estamos a falar normalmente de recomendações que têm uma forte componente política. Ou seja, grande parte das recomendações que acabam por vir das avaliações que fazemos, sobretudo

nestas avaliações mais de caráter de impacto ou mais estratégicas, é uma componente de depois de decisão associadas a elas, são fundamentalmente decisão política, não são meras de dimensão técnica. (...). Até porque também as próprias recomendações que vem nas avaliações, no follow-up, há recomendações que decidem não adotar. Mas queremos uma justificação, não vamos adotar por esta ou aquela razão, que há outros motivos que às vezes até são discutíveis, que os avaliadores não tiveram em conta e não seguimos essa recomendação, não deve ser adotada, pelo menos não deve ser seguida tal qual está a ser a proposta porque temos estas ou estas razões. E, portanto, isso é aquilo que decorre no próprio processo follow-up, depois fazemos as conclusões e as recomendações que os avaliadores produzem.

Voltando só um pouco atrás, escapou uma questão: como é que, no caso desta avaliação que foi feita, como é que foram equacionadas e definidas realmente estas abordagens metodológicas, ou seja, porquê estas e não outras? Falarem-nos, se puderem, sobre a escolha do caminho metodológico.

P1: A escolha do caminho lógico partiu de 2 propostos, sendo estas tipicamente avaliações focadas na avaliação de impacto, ou seja, qual é o contributo dos programas para aquilo que é o desempenho ao nível de alguns indicadores de resultado. Por exemplo, eu tenho um indicador de desenvolvimento, que é a empregabilidade dos jovens, vou verificar 6 meses após se os formandos que eu apoiei concluíram o curso e estão empregados ou não empregados, vejo que 50% ou 60%, 70% sim; isso é bom, é mau? não sei! (...). É preciso ver que há um grupo em concreto e comparar com quem não fez (...). É este tipo de exercício que se quer. Procurar no fundo que essas avaliações forneçam o máximo de informação possível para perceber qual é que é o contributo efetivo dos programas. (...).

E daquilo que foi efetivamente depois a avaliação, ela conseguiu efetivamente dar resposta às questões fundamentais e aquilo que eram as expectativas que tinham sobre a avaliação?

P1: Eu diria que globalmente sim (...). [E] os exercícios de avaliação que, em regra geral, foram coordenados nesta área eram, em regra geral, muito exigentes e amplos, ambos, quer pela dimensão dos atores, quer pela dimensão de programas e tipologias de intervenção que abrangiam. Isso complexifica mt as avaliações, mas acho que, globalmente, acho que o resultado é positivo. Sinal disso mesmo é que nós temos vindo a utilizar esses resultados para apoiar e sustentar o processo de programação – apesar de algumas críticas ou preocupações que a Comissão Europeia tem manifestado em relação ao nosso modelo de avaliação, porque os resultados chegavam mais tarde do que queriam, etc. Mas, ainda assim, no final do dia, conseguimos chegar a esta fase do processo de programação e com um conjunto de exercícios de avaliação que deram resultados, que foram apresentados, foram discutidos, e estão a ser utilizados para sustentar e apoiar períodos de programação. Eu diria que o aumento do balanço é positivo, independentemente de haver sempre coisas a melhorar. Temos feito o nosso caminho e o nosso esforço nesse sentido. (...). Nós temos usado muito essa avaliação para fundamentar a programação do PT2030, em particular naquilo que tem a ver com os objetivos do fundo europeu, (...), para fundamentar as suas opções de programação.

Pensando nas equipas de avaliação, o que é que poderia ser melhorado, ou que lacunas é que encontraram naquele processo de avaliação, que foi feito há uns anos atrás, que nos permitam aqui de alguma forma, também, neste modelo que nós procuramos aperfeiçoar, ter isso em linha de conta?

P1: Uma das dificuldades que nós temos, também com alguns progressos, mas nem sempre conseguimos ir tão longe como isso, é exatamente que, por um lado, as conclusões, às vezes, temos aqui uma dificuldade em que elas estejam tão sustentadas quanto nós gostaríamos nas evidências da avaliação. Temos tido aqui alguma discussão, por vezes, com as equipas avaliação, porque às vezes, acabam por cair na tentação, porque a sua equipa já é muito experiente, já conhece muito bem o objeto que estão a avaliar, e às vezes também já vêm com pré-conceitos sobre as coisas e, portanto, às vezes há alguma dificuldade em desligar disso. Ou, às vezes, algumas evidências serem sustentadas em evidências não tão sustentáveis quanto isso... E depois tem implicações na questão das recomendações, não é? Nós precisamos às vezes de recomendações muito mais operacionais, até para facilitar o processo de follow-up, e temos recomendações muito genéricas, que são muito difíceis, na prática, de operacionalizar e de se traduzir em coisas mais concretas em matéria de produção das políticas públicas que foram objeto de avaliação.

P1: Nas conclusões, às vezes temos tido algumas dificuldades, porque não nos apercebemos logo e saem coisas que refletem quase a opinião, pelo menos, de alguns stakeholders que foram ouvidos e sem grande filtro nisso; ou às vezes, também, dos próprios avaliadores, que já são muito experientes, já têm muitos estudos anteriores e estão a reproduzir um bocadinho o que já têm no seu mindset, às vezes quase independentemente dos resultados da avaliação que fizeram. E depois, ligado a isso, a questão das recomendações serem o mais objetivas e mais pragmáticas possíveis, para facilitar, depois, o follow-up.

Se lhe perguntasse agora de forma oposta, aspetos que considera que tenham sido mais bem concretizados nesta avaliação que foi feita, ou mesmo noutras... aspetos mais relevantes?

P1: Para isso também contribuiu a componente contrafactual, mas não só. Porque, nos precisamos de avaliações que saibam comunicar os resultados. Como sabem, acho que às vezes, quase que há um preconceito contra aquilo que é o investimento dos fundos europeus, ou por via das questões da fraude - sempre que aparece um caso de fraude, isso tem um impacto imediato e muito grave. E depois, também, do ponto de vista da evidência dos resultados, de se combater alguma desinformação que ainda persiste sobre isso. É importante mostrar aquilo que, sem os Fundos, não teria sido possível fazer. Acho que esse é o grande valor acrescentado que as avaliações nos têm trazido e que nós temos procurado: contribuir para um debate mais qualificado, conhecendo o que está menos bem feito, o que é preciso fazer melhor? Mas também reconhecendo que aquilo que foi feito, contribuiu para mudanças relevantes e estruturais nas áreas de intervenção que estamos a falar.

Uma questão, referiu a importância do contrafactual, sendo que o contrafactual se limitou a uma questão, uma área do desempenho académico. Existem outras questões ou áreas da avaliação que tenham sido menos concretizadas, com menos capacidade de concretização devido a não ter sido realizado um contrafactual, por exemplo?

P1: Provavelmente sim, eu diria que sim. Agora há uma dificuldade: como estamos a falar de áreas muito diversificadas, encontrar o contra factual para cada uma daquelas áreas, nalguns casos, não sei se era um exercício que se quer...

P2: não sei sequer se temos pessoas suficientes dentro das amostras.

P1: Normalmente, nos contrafactuais, existem dimensões da mostra, ter um grupo comparável que permita fazer esse contra factual, e às vezes ou é difícil às vezes pode mesmo não existir. Vou dar um exemplo, nós financiamos formação contínua de docentes; fazer um contra factual sobre isso, a questão é que provavelmente, pelo menos nas regiões em que nós intervimos, toda a formação foi feita, e abrangeu um grande universo, foi feita com Fundos Estruturais. Portanto, a maioria fez essa formação, talvez só ficou de fora aqueles que já não quiseram ir. Ou seja, a maioria, pelo menos de uma ação beneficiaram. Nestas áreas de intervenção tenho dificuldade em fazer um contra factual, porque não tenho quase áreas de intervenção em que não tenha tido intervenção, neste caso, os dos fundos ou outras fontes, não é? Portanto, é difícil fazer esse contrafactual, pelo menos com os níveis de exigência metodológica que as políticas dos Fundos Estruturais exigem.

Mas depois sente a tal dificuldade com a comunicação dos resultados e dos benefícios das medidas?

P2: Aqui a grande vantagem, é nós podermos utilizar metodologias complementares e essa. Naquelas em que a metodologia não permite chegar a um conhecimento, neste caso, ao contrafactual, não permite não permite chegar aos efeitos de ter sido tratado ou não tratado, é porque temos a outra componente, neste caso da teoria da mudança que nos permite chegar a outro tipo de conclusão. Porque o que é preciso, e o desafio é esse, da parte das equipas de avaliação, é conseguir montar o puzzle, de maneira que se perceba que as opções metodológicas utilizadas são mais efetivas para uma componente e menos para outra e, portanto, aquilo que tem de dar um resultado de que aquilo é coerente do ponto de vista da comunicação, que era como P1 estava a dizer anteriormente.

P2: Portanto, no fim, no fim de contas, as conclusões a que chegamos tem que ser muito bem fundamentadas de acordo com as opções metodológicas que se tiveram subjacentes a todo o processo avaliativo, esse é que é o grande desafio que as equipas de avaliação têm de ter: perceber que nem tudo é possível fazer, porque não há, não é mesmo realista. Este exemplo que P1 estava a dar é muito bom: se nós não temos ninguém que fique de fora do grupo dos tratados, não há contra factual que seja possível realizar. Por isso, nós optámos por fazer contra factual dentro daquela intervenção que mais fazia sentido utilizar: no

caso era os cursos profissionais, até porque o peso em termos de representatividade na população estudantil em Portugal, até permite fazê-lo, ao passo que nas outras áreas não é possível. Mesmo que se fosse possível fazer um contra factual, não ia ter representatividade nenhuma, portanto, tinha pouco valor em termos científicos. Não valia a pena estar a desencadear um processo contra factual que não ia ter valor, não ia permitir ter conclusões que não fossem fidedignas. Agora o desafio depois é pegar em todas as opções e montar conclusões e depois recomendações. E que depois nós tiremos recomendações que permitam, a partir daí, nós tomarmos decisões. O valor da avaliação, para além de perceber aquilo que foi feito, é o que é que nós podemos fazer a seguir! O desafio é termos recomendações que sejam operacionalizáveis para a política. As coisas têm de ser coerentes.

Podemos, de alguma forma, considerar que esta é uma das principais expectativas que têm, então, em relação às avaliações, é este contributo?

P1: Sim. Tem esta dupla vertente: o que foi feito, de que forma foi feito, e o que é preciso fazer diferente, ou igual, e a partir daí, o que é que vamos decidir a seguir.

A última questão: se tivessem oportunidade de mudar alguma coisa na avaliação que foi feita, se alteravam e o que é que alteravam?

P1: Eu penso que se houvesse alguma coisa que eventualmente se alterava, é que a avaliação em causa, tinha um objeto muito amplo, muito diversificado, abrangia uma grande diversidade de intervenções e, portanto, isso teve a vantagem de ter a visão uma visão geral, de conjunto, mas ao mesmo não permitiu, para algumas áreas específicas de intervenção, ter o nível de profundidade que se exigiria para algumas intervenções, até algumas delas com uma relevância bastante grande, para o período de programação. São contributivas para aqueles objetivos, mas que ficaram bocadinho, porventura, menos cobertas pelo próprio exercício da avaliação. Mas isto também foi tentar um equilíbrio entre o número de avaliações mais global, mais transversal, mas com essa transversalidade, em algumas áreas não permitiu aprofundar muito esses resultados. É uma coisa que agora estamos a procurar compensar ou reforçar agora, com as avaliações intercalares (...).

P1: Uma das conclusões da avaliação foi que estas políticas produzem resultados que fazem parte de um ecossistema de políticas, algumas delas elas nem sequer financiadas pelos fundos europeus, mas que no fundo convergem todas no mesmo sentido, e com o mesmo foco, por isso estão a produzir resultados.

Podemos achar que no lugar de um caderno de encargos podiam ter havido mais, de forma, então, a segmentar um bocadinho e poder aprofundar, com prejuízo para uma perspetiva mais geral? Seria uma alternativa?

Sim, no limite sim. (...).

Sabemos que houve uma avaliação ex ante e aquilo que gostávamos de saber era que expectativas tinham em relação a esse exercício? E de que forma é que ele poderia depois, então, contribuir para o próprio desenho do programa?

Pessoa entrevistada refere que não acompanhou essa avaliação: “Não acompanhei a avaliação e não fui utilizadora dos seus resultados, porque essa avaliação serviu supostamente para sustentar a programação; não é para fazer avaliação nenhuma, é para sustentar – aliás, é uma obrigação, há aqui todo um circuito obrigatório, que é relativamente rígido. Muitas vezes traz poucos contributos depois reais. São exercícios mandatórios e, portanto, servem para sustentar e para preencher determinados pressupostos. E depois, na fase em que as avaliações ex ante são feitas – agora já estou com uma posição crítica e não propriamente de Secretária técnica – na conhece-se muito pouco – e pensando agora na teoria da mudança – conhece-se muito pouco ainda quais é que vão ser as os grandes objetivos dos próximos períodos de programação.

(...) aquilo que é o pressuposto na avaliação ex ante – e pensando nos Fundos, como as coisas estão sempre muito atrasadas do ponto de vista do planeamento, em termos de calendário, e pouco densificadas em termos de grandes objetivos e de grandes metas, torna-se inviável que os resultados de uma avaliação ex ante sejam de facto, depois, positivos para o desenho dos programas. Há um ciclo de suposta utilidade, uma cadeia de valor – porque estas coisas todas começam a ter que ser faladas na lógica de uma cadeia de valor – que não se dá. Por isso é que eu digo que são muito mais exercícios rígidos e

mandatórios do que propriamente úteis para a própria cadeia de valor do planeamento nos Fundos, para a planificação. Há um desfasamento temporal; ainda não há muita coisa preparada e planeada para quem a faz poder utilizar, para poder ser útil para a fase seguinte. E qual é que é a fase seguinte? É que o planeamento do programa, o detalhe do programa, tenha em conta os resultados dessa avaliação.

A ideia que tem é, no fundo, sobre suportar; o que é que é exatamente suportar?

Os resultados da ex-ante suportarem o desenho específico do programa.

Isso supõe que havia um desenho anterior ou que na ex-ante é que o faz? ...nessa altura também não é claro qual é o desenho que vai ter o programa, porque os objetivos eram estabelecidos, porque ainda não houve tempo para quem vai desenhar ter um verdadeiro diagnóstico, ou consultar alguns estudos, alguma ideia sobre quais são as medidas que devem ou não ser contempladas, ou o quê?

Quando se olha para o detalhe, para o que são os elementos que configuram o programa, olhas para quais são os elementos core que definem o que vai ser o programa, que são as gavetas financeiras alocadas, em função de determinadas metas. Para se definir isso de forma supostamente coerente, portanto, com todas as dimensões que depois configuram questões de avaliação mais tarde, ou seja, serem eficientes, eficazes e por aí fora, tu tens que ter determinadas orientações e confirmações prévias que aquilo vai resultar e que não seja apenas o histórico, porque em determinados programas há uma reprodução – aquilo funcionou, continuamos com necessidades que justificam que se volte a alocar financiamento para isto, mas a verdade é que houve uma evolução, porque senão os programas anteriores e os ciclos anteriores não tinham feito não tinham feito nada.

E o que eu acho – mas notem, eu não vi os resultados da avaliação ex ante, não conheço os detalhes deste programa – mas se estamos exatamente a financiar da mesma maneira por exemplo os cursos profissionais que estávamos a querer financiar em 2013 estamos mal. Porque em 2020, a taxa de abandono escolar precoce, que era a grande justificação dos cursos profissionais, já não tem nada a ver com o que era em 2012 ou em 2013. E o próprio gestor do programa passou o ciclo a dizer que agora o que nós temos de fazer, relativamente à taxa que ainda sobra, é ir às franjas. Isto é uma coisa que tem que estar numa avaliação ex ante. Portanto, se o problema já é diferente, como é que o programa vai exatamente alocar o mesmo valor, vai exatamente dirigir-se ao mesmo universo, vai exatamente querer atingir as mesmas metas? não percebo.

Parece-lhe, assim, que a avaliação ex ante forma ou não as opções que depois são tomadas por quem desenha o programa? Ou é uma coisa que depois é variável: depende de quem é, depende das circunstâncias, depende do programa?

Acho que justifica o Programa, serve para legitimar oficialmente.

Há outro aspeto que surge bastante, que é a ideia da não explicitação da teoria da mudança. Mas há avaliações ex ante e que deviam, precisamente, perceber a lógica com que o programa foi desenhado. Então, se estamos a perceber como é que o programa foi desenhado, qual foi o diagnóstico, quais as necessidades, quais foram os recursos a alocar, porque é que escolhemos estas medidas e não outras para responder a estes problemas, então devia estar explicitado. Mas depois dizem que não, que não houve explicitação. Parece contraditório. Há uma execução, da avaliação ex ante, mas depois não é considerada.

A principal medida que esteve em avaliação na avaliação do POCH é muitíssimo expressiva do ponto de vista orçamental para Portugal, é muito expressiva. Se aquela fonte deixasse de existir, o orçamento de Estado teria que desembolsar +2.000 milhões... Acho que isto determina muitas conclusões quando se faz avaliação deste tipo de medidas a serem financiadas pelos Fundos, porque tem imenso impacto financeiro. Acho que é isto que de certa forma torna menos útil e condiciona certos exercícios. Acaba por não se investir tanto nas reais utilidades dos exercícios de avaliação, porque há um objetivo muito maior, que é político: o financiamento dos cursos profissionais é uma necessidade orçamental e uma dependência muito grande dos Fundos.

Uma questão sobre o diagnóstico: há um real conhecimento dos problemas e dos fatores que contribuam para os problemas, e da evolução da situação, ou se até há esse conhecimento, mas é na transição para as medidas responderem a esse problema que está a dificuldade? Como lidar, por vezes, com diagnósticos desatualizados?

No caso dos problemas da educação, os problemas têm vindo a evoluir e acho que há uma noção clara desse diagnóstico. Outra coisa é, se aquilo que ainda resta resolver, E volta a taxa de abandono escolar precoce, aquilo que está por resolver é passível de ser resolvido com o título de medidas que os fundos projetam? Eu acho que isto é outra coisa, outra pergunta. Porque no que respeita à educação, os fundos são apenas um instrumento de política, não há exatamente fundos para a educação. (...). Os fundos fazem mudanças estruturais na política de educação? em si, não fazem. Em si, são apenas uma fonte de financiamento. Sem os fundos e se podia vir por outro meio qualquer, por outro canal qualquer. Estarmos simplesmente a financiar os cursos profissionais, não vejo como é que isso vai resolver os 6% ou que seja de taxa de abandono escolar precoce.

Há um desencontro entre quem desenha as medidas e quem as aplica... Mas quem está nas medidas, os stakeholders, estão nos grupos de acompanhamento dos programas... Para além disto, o estatuto que têm na avaliação é de acompanhamento da avaliação; mas em que medida é que estão comprometidos com a avaliação, se não são considerados no desenho?

Sim, o desenho do programa é político.

Na participação que teve no Programa e na avaliação, tem ideia de que, de alguma forma, os resultados da ex-ante possam ter sido trazidos para o Programa, ainda que de forma breve?

Não tenho. A partir do momento em que o programa é definido aquilo que passa a ser a luz, perdem-se as outras referências. Embora haja períodos em que seja suposto haver reprogramação, mas as reprogramações não são de fundo, não são de princípio, são de estratégia, mudar ou afinar coisas. Não se volta de princípio. Estas reprogramações são feitas com base nos níveis de execução, é um processo de gestão normal, mas em vez de estarmos perante um projeto, estamos perante um programa, que tem várias linhas de ação e dentro de cada uma dessas linhas, faz-se um Balanço de como é que se está a meio do processo.

Que expectativas é que tinham em relação a esta avaliação?

ADC, enquanto coordenadora de princípios gerais, formas e linhas de ação comuns dos vários programas, considerou, e bem, que a avaliação devia ganhar um novo élan. acho que todos percebemos isso em 2015, 2016, no arranque do anterior ciclo: vamos dar um grande enfoque à avaliação, a avaliação tem de passar a ser útil, tem de ser bem vista. Havia essa estratégia. E para isso munuiu-se de muitas orientações: como fazer? O que vai ser? Como todos temos de contribuir para isto? Muitas regras. O Plano Global de Avaliação, a tal bíblia de todos os programas, tem detalhes impressionantes.

Quando entro, em outubro de 2016, tenho um plano em que estão previstas cerca de 50 avaliações e há uma ficha para cada uma delas. E essa ficha já me diz qual vai ser o método de avaliação, vai ser uma avaliação do processo, se vai ser uma avaliação de impacto, já diz uma série de coisas. já diz quais vão ser as questões de avaliação, quanto vai custar a avaliação, quem são os organismos que vão participar, portanto, tem BI todo da avaliação. Claro que isto para programas menos capacitados é ótimo, é altamente agilizador porque tens já um bom calendário, é mesmo muito bom. (...). Eu tinha um plano para seguir, e por exemplo, na ficha estava escrito que eu ia fazer uma análise contra factual, e eu faço não posso sair dali nem pôr em causa se é o melhor ou não. Havia todo um arsenal de como fazer, que era muito útil, mas muito apertado. (...).

Um dos problemas que surgiu teve a ver com o confronto com o mercado, ou um não mercado, com sermos altamente rígidos nas especificações técnicas. Eu lembro-me que não eram os programas que tinham definido qual era o método, eu lembro-me quem determinava e propunha quais seriam as melhores metodologias para responder àquelas perguntas era quem sabia de avaliação. A questão é que quando ADC tem esta visão, ela está a trabalhar com 2 pessoas ou 3 que percebem de avaliação, mas elas não têm que perceber de avaliação. ADC não tem de perceber de avaliação como um especialista em avaliação, tem que possibilitar e integrar processos avaliativos no ciclo dos fundos, é só isso que tinham de fazer. Acho que criaram demasiadas regras ex-ante para o mercado, demasiadas orientações para o mercado. E o mercado sentiu-se completamente num espartilho, com muita pouca liberdade científica para olhar para os processos e essa pouca liberdade cria um conflito de

papéis. (...). Mas há uma nota de rodapé, que acho que não é uma nota de rodapé, que é científica, e que os resultados da avaliação do POCH São apenas representativos de 30% do universo que um programa financia.

Então só uma pergunta de confirmação: na verdade as opções sobre as metodologias e sobre como a avaliação deveria ser feita decorre muito mais de opções nacionais do que de diretivas europeias de elementos que devem constar nas avaliações dos programas.

Foram opções nacionais levadas à letra, ou seja, foram demasiado rígidos. Na Comissão Europeia começaram a dizer que nós, Fundos, Temos que começar a ser como os tipos do Banco mundial e temos de começar a provar que o dinheiro que damos serve para mudar qualquer coisa. E isso faz-se muito bem recorrendo ao contrafactual. Esta é a minha leitura. Mas há vida para além do contrafactual. O contrafactual aplica-se a coisas que têm características específicas. E quis-se enfiar e justificar que era possível aplicar essa técnica a determinadas coisas que de facto não têm, ou que têm muito pouco, e que põe em causa os resultados que se obtiveram. Vamos olhar para as bases de dados da DGRegio, E comparam os exercícios de Portugal com os de os outros países todos e vê lá onde é que está o contrafactual, a teoria da mudança, e quem é que seguiu aquela linha de pensamento da Comissão Europeia? Ninguém. Mas nós, Portugal, hiper bom aluno, by-the-book, 'bora lá. O que importa é ter muitos guias, muitos guias, e é um exagero. E depois ainda se dá a coisa mais caricata que é: nós estamos a fazer exatamente aquilo que a Comissão orientou os Estados Membros a fazerem, mas se somos tão bons alunos que fazemos isto num ano e meio e a Comissão diz que não, que é para fazer em 3 meses. Porquê? Porque eles têm matrizes do que são os projetos e os resultados de avaliações do Banco Mundial. que se fazem 3 meses, que são as avaliações rápidas. Mas isto não se faz com o tipo de projetos que os Fundos apoiam.

É importante olhar para o plano e para a cultura daquilo porque o plano é muito diretivo. Foi desenhado por ADC, (...). Acho que os métodos e as técnicas não tinham de estar definidos à partida. Acho que o cliente tem de dizer o que é que quer ver respondido, mas devem ser os avaliadores os potenciais avaliadores a dizer como é que se chega às respostas. Nós passamos muito tempo com a equipa dos jovens a discutir a pertinência do contrafactual, e com visões contrárias. É útil uma equipa de avaliação estar a trabalhar contra vontade? Não é, não vai produzir o melhor, por mais séria que seja não vai produzir o melhor.

Havia este plano geral de avaliação e uma ficha para o POCH. Ainda que a estrutura da ficha não vos parecesse o ideal, porque era muito fechada, de alguma forma conseguiram negociar? A estrutura do POCH interveio no caderno de encargos e no resultado final daquilo que foi para o concurso?

A questão é que os atores são os mesmos, quem desenhou as fichas do POCH também foi o gestor do POCH. São atores que são os mesmos, é uma rede pequena. Mas aquilo que tinha de ser feito no âmbito da ADC, no âmbito do plano? Se calhar não tinha, mas o pensamento foi: vamos já fazer porque a malta dos programas não tem competência para fazer isto, então vamos já orientá-los para fazer isto. É um trade off. Eu acho que isto tinha de ser analisado, os prós e os contras disto. Para mim isto é uma questão importante: como é que se lida com demasiadas orientações, e centralizadas, em benefício de haver pouca capacidade dos programas, com um espartilho que retira ao mercado um interesse e uma procura desejáveis?

Da sua experiência/perceção sobre a avaliação, as abordagens metodológicas têm limites, são boas para aferir determinadas dimensões do problema ou da avaliação, mas depois têm limites para outras coisas. Estou a pensar nos cadernos de encargos e nas avaliações e pergunto 2 questões: primeiro se o entendimento das abordagens metodológicas é o mesmo entre quem está a pedir e quem vai depois candidatar-se – são abordagens metodológicas, se vamos ver teorias da mudança, por exemplo, diferentes autores têm diferentes perspetivas e isso faz com que considerem que algumas coisas estão englobadas e outras não, e que permite aferir determinados resultados ou não; e também perguntando da parte de quem encomenda – porque estamos a discutir nesta parte a definição do modelo de avaliação – que é, se depois, o que estão a pedir – porque uma coisa é dizerem que querem fazer isto com a TdM, outra coisa é dizer que querem testar a eficiência, se a TdM é um bom instrumento para responder a isso. Ou seja, se entre a abordagem metodológica que se pretende e os resultados que se querem, se há um casamento, se faz sentido, se é coerente, e se quando é feito o primeiro relatório, em que se apresenta o modelo, em que se valida o modelo de avaliação, se as pessoas estão de acordo sobre os limites e as potencialidades dos instrumentos?

No lugar em que eu estava... quanto ao POCH e o contrafactual isso percebeu-se os limites. mas tenho um exemplo de um contrafactual muitíssimo bem feito no sentido em que a razão para se estar a usar o contrafactual é a correta e os resultados a que se chegaram decorrem precisamente disso - os resultados não é de serem bons ou maus para a política, mas o tipo de resultados que ali estavam.

A questão é, durante a avaliação houve alguns momentos em que dissesse que se queria testar, por exemplo, a eficiência e houve dúvidas sobre se a abordagem metodológica para esta avaliação era um bom instrumento para testar a eficiência?

O cliente não tem de discutir isto. Eu podia não perceber nada de avaliação. Eu só tenho de executar um plano. Eu só tenho de garantir no caderno de encargos que faço é compreendido, tem boas condições para ter várias propostas, que coloquei lá as perguntas que são relevantes para a avaliação de uma medida, de um programa, de um eixo, do que seja. Acho que esse tipo de coisas nós não nos devíamos preocupar, porque repara, se esta avaliação estivesse a ser a dos jovens, se estivesse a ser conduzida do lado de um PO que não tivesse pessoas que percebam um pouco do assunto, a equipa formulada pelo Luís tinha convencido, e bem, que não tinha de ser feito o contrafactual, que não podia ser feito contrafactual. Mas o nosso gestor dizia que não, nós queremos que seja feito, já foi feito e queremos que seja feito novamente. Tem limitações sim, mas nós assumimos essas limitações e queremos que isso seja feito. Talvez se fosse outro gestor, que nunca estivesse estado ligada à avaliação, talvez houvesse abertura para ver o que é que podia ser feito de forma diferente. Eu acho que aqui o científico ainda tem superioridade relativamente a outras coisas.

Proponho voltarmos um pouco à teoria da mudança

Eu tenho a dizer: sempre se fez teoria da mudança, mas sem esse nome, e o que se atinge em termos de avaliação, em termos de resultados, com esse nome, é o que já se atingia antigamente. Antigamente chamava-se quadro de análise, essas coisas todas... Hipóteses? Não percebo porque é que há hipóteses. (...). Mas quais hipóteses? Eu olho para os resultados daquilo tudo e para a sinalização daquilo tudo, e não acho que gere nada de novo, é muito confuso, gera esquemas altamente confusos, mais cheios, as lógicas, os mecanismos e não sei quê... Numa avaliação como a do eixo 1, como a dos jovens, que tinha mais de 6 tipos de medidas, desde os cursos profissionais até ao PNPSE, não faz sentido é uma loucura. Até interessante do ponto de vista do passado do pensamento, tudo bem, mas aquilo não gera outputs diferentes do que as anteriores avaliação geravam.

Não traz, portanto, nenhum acréscimo? Que acréscimo é que pode trazer? Aquilo que me parece é que complexifica talvez demasiado, para depois mais tarde, o resultado não ser mais vantajoso...

acho que tem uma vantagem de, é um guião é um roteiro para não perdermos de vista que aquilo que vamos ter de responder no fim é aquela pergunta do: isto foi feito para quê? Mas isso, no antigamente era a pergunta de partida, que tem em conta os objetivos. Quem não sabe estas coisas – quem faz a avaliação é quem vem das ciências sociais, sabe isto. quem é que criou isto da teoria da mudança? Não fui investigar, mas tenho a sensação de que há aqui um baralhar e dar de novo, muito típico das teorias de síntese, e querer dar outro nome a coisas que já têm outro nome mas estão noutro campo científico, não sei... (...) Trata-se de títulos, de nomes, porque o tratamento é idêntico.

Do vosso ponto de vista, perante a situação, qual é que consideravam que poderia ser o contributo, ou o papel da teoria da mudança nessa avaliação? Admitindo que ela ia estar presente, qual era a vossa expectativa?

Não tinha expectativa. Quer dizer, tinha expectativa de encontrar coisas diferentes nos resultados. Eu acho que a teoria da mudança é um modelo analítico para processos de planeamento, para projetos, para planificação, quer-se fazer isto porquê, porque agnóstico, por porque temos estas medidas, com estas medidas queremos atingir estes resultados... Isto é um processo de análise para processos de planeamento.

Como dizia, é um modelo analítico, tem o porquê, tem o como... há uma lógica, há uma intenção de tentar perceber o processo desenho, planeamento e execução de uma política pública nos termos em que ela foi pensada. Isto tem pressuposta a ideia de que fazemos isto para que as pessoas aprendam com o que fizeram. (...). A pergunta que eu faço, é se tem a sensação de que no final as pessoas só retiram os resultados a que a política chegou, ou se realmente aprenderam alguma coisa com este processo, seja para um novo processo de planeamento, seja para ajustar medidas?

Eu já não acompanhei o ciclo do desenho, (...) mas eu diria que se lá estivesse – que nós fazemos uma leitura de falou das avaliações, que são dados muito importantes, o que é também um processo complicado, mas isso seria outra discussão... Eu se estivesse a fazer o desenho agora do novo ciclo, portanto teria de voltar a ler os resultados da avaliação nessa perspetiva, para encontrar aquilo que me diz quisto deve ser feito e que isto não deve ser feito, ou seja, usava aquilo com recurso. E antes de pensar o que quer que seja, porque aquilo também faz parte do diagnóstico, aquilo integra a fase zero de um processo de planeamento.

Há uma coisa que vale a pena dizer: os esquemas da teoria da mudança, para quem não está por dentro da teoria da mudança, não percebe nada. (...). A questão é, qual é o pensamento por detrás das medidas, por exemplo, dos TEIP, dos cursos profissionais? Isso não estava muito explícito, mas depois quando se foi ao terreno, quando se fez a avaliação, isto apareceu. Será que para quem geria as medidas E fazia parte da comissão de acompanhamento para avaliação tem a mesma utilidade, ou não?

Não, a avaliação vista pelos outros é um exercício do programa.

E o programa financia, mas não pode fazer nenhuma recomendações?

No processo follow-up, diria que foi muito mais importante do que a teoria da mudança. Aquilo que acho que rompe com o passado é ter-se instituído o processo de follow-up das avaliações. Isto obrigou a termos de, no programa, bater à porta de muita gente, para dizer que a avaliação X recomendou que vocês façam isto, isto e isto, digam-nos se vão fazer, sim ou não, se vão fazer, quando é que vão fazer, e que plano têm para o fazer. Havia a hipótese de não fazer, mas refletindo sobre isso, ou seja, porquê que não querem implementar o que foi recomendado. Para mim isto foi a coisa mais importante que foi implementada por ADC, e que vai demorar muito tempo a produzir resultados positivos. A introdução do follow-up fez parte do pacote de orientações e de guias que ADC produziu. Há um guia específico para o processo de follow-up, cheio de fichas e etc. Eu dei os parabéns a quem visionou o modelo que estava nas fichas, mas também disse que o que ali estava, ia exigir muita conversa, muita cultura, muita imersão, não se vai passar a dar importância à utilidade às avaliações porque há fichas para responder. E isto bate naquilo que o Alexandre estava a perguntar: é útil para os stakeholders? Eu diria que sim, porque aqui estamos perante avaliações que são ao mesmo tempo avaliações dos programas, mas são, no caso do POCH, avaliação de políticas públicas portuguesas. Outros não, outras medidas como por exemplo a inovação social, só existe no contexto dos Fundos. No caso de tudo aquilo que faz parte do sistema educativo, são medidas que sim, deviam ser úteis para esses dirigentes todos que executam essas medidas e que não têm outros processos avaliativos, porque a avaliação também custa e eles priorizam investir noutras coisas.

A questão da monitorização do programa: havia um plano e de que forma é que esse do ano contribuiu para o processo de avaliação, e que contributo é que os vários stakeholder se deram para esse processo de monitorização?

Os stakeholders não contribuíram para o processo de monitorização. A monitorização é uma exigência que está parametrizada nos regulamentos da Comissão. Para quê que foi útil? Para o nosso diagnóstico, ou seja, o caderno de encargos estava enriquecido com o estado da arte do programa. E isso foi feito com dados da monitorização. Aquilo que antigamente acontecia nos processos de avaliação era que as equipas ainda demoravam mais tempo a entrar mesmo na avaliação, porque tinham de ter à sua frente dados sobre como estava o programa. Isso já está feito agora, já vem feito. Quem começa a fazer a avaliação já têm à sua frente quanto é que se gastou, quantas pessoas foram abrangidas, onde é que se está relativamente às metas, já tem algumas pistas para as primeiras questões, como por exemplo eficiência, eficácia, etc. Isso é feito com estes elementos.

As entidades beneficiárias do POCH, elas também são forçadas a ter um sistema de monitorização para terem um financiamento?

Sim, têm de tê-lo para prestar contas. É um efeito/modelo cascata. A Comissão tem, os Programas têm e os beneficiários têm.

Ainda do ponto de vista metodológico, falamos já muito sobre a teoria da mudança, podemos falar agora um pouco mais sobre o exercício da análise contrafactual? Pese embora toda a orientação do caderno de encargos, o facto já estar fechado, que expectativas que tinham, como é que correu?

As questões do acesso a dados foi uma das principais vicissitudes desta avaliação. o acesso a dados administrativos, que é a matéria-prima de um exercício contra factual, é cada vez um processo mais complicado e, portanto, é um fator de risco muito grande para o sucesso de uma avaliação. Depois como entrámos em confinamentos ainda foi mais difícil, porque o organismo que alimentava essas fontes exigia que as pessoas tivessem lá, mas depois não podiam estar lá, e fez com que o processo parasse algumas vezes. Este é um risco muito grande as avaliações em Portugal. E o contrafactual vive deste tipo de dados, já as entrevistas à distância correram muito bem, gasta-se menos dinheiro, o tempo para essas tarefas é menor, portanto ganha-se tempo. As avaliações em Portugal deviam ter uma via verde para os dados administrativos, não se fazem avaliações robustas, claras, sem dados administrativos. E Portugal, daquilo que eu tenho ouvido, é incrível em dados administrativos: nós temos dados para tudo.

Gostaria de ter sabido os resultados do exercício alternativo que a equipa de avaliação propôs, quando o exercício do contra factual caiu. Era uma abordagem muito diferente, gostava de saber que resultados íamos alcançar, era relacionada com a questão das desigualdades...

Considera que esta avaliação conseguiu responder àquelas que eram as principais questões colocadas e se houve áreas que considera que tenham sido mais bem concretizadas, outras que pudessem ter sido de outro modo, algumas lacunas? No fundo um Balanço...

Há uma pergunta que está na moda que é o valor acrescentado, o VAE. eu ainda estou para ver uma avaliação que tenha coisas relevantes nesta dimensão, porque quando digo isto, estou a pensar e os avaliadores têm alguma dificuldade em não dizer coisas redundantes nesta dimensão da avaliação. Ela é obrigatória da parte da Comissão, aliás, é a bandeira da avaliação na medida em que os justifica, justifica todo este modelo de financiamento dos Estados membros através dos Fundos, porque se os Fundos não existissem, ou por outra, em que é que os fundos contribuem, qual é o valor adicional para esta política existindo os fundos? Isto vai de encontro ao que dizia há pouco, para mim, é a nossa balança orçamental ficar um pouco mais aliviada. Isto não se pode dizer, com certeza. E depois a avaliação como comportava medidas muito distintas entre si - e eu sobre algumas não tinha qualquer opinião, qualquer evidência do resultado que produziam - A ligação foi muito importante para dar algumas pistas, sublinho, pistas. Porque sobre isto estamos a falar de políticas, e eu acho que se confunde ali coisas por exemplo não é pelo número de psicólogos que se sabe se aquilo tem ou não efeitos positivos nos resultados, no sucesso dos miúdos. Aquilo que se fez foi uma monstra avaliação, é um monstro, e há coisas que se perdem. Centrou-se muito na questão dos jovens, e depois as outras medidas, que eu acho que são muito importantes, as tais que vão fazer diferença na história da taxa de abandono escolar precoce, já não vão ser os cursos profissionais, aí acho que já não vão fazer nada... Não sei se há avaliações suficientemente robustas que produzam as devidas evidências para manter essa política a ser cofinanciada.

Mas esses resultados são porque eu não tenho a certeza de que aquilo tenha sido avaliado com tempo, com recursos, com as melhores abordagens, porque os cursos profissionais são um pouco Como Eu os eucaliptos: secaram as outras áreas. Eu hoje batia-me por não permitir uma avaliação daquele tamanho, daquelas medidas todas. Deveriam ser vários cadernos de encargos, porque por exemplo, gastar o dinheiro para avaliar os psicólogos daquela maneira?

Tem a sensação que a Comissão Europeia tem ou não noção de que está a financiar medidas mainstream? Eu penso que eles sabem, eles não são assim tão ingénuos. Acho que têm a noção de que de alguma forma sabem que estão a financiar medidas de política nacional...

Sim, eles sabem. Existe uma categorização em que nós estamos juntamente aos outros países que foram entrando, mais recentemente, em que temos um alto nível de dependência. Sabem que sem aquilo os nossos níveis de desenvolvimento eram de facto menores, porque estamos a financiar coisas estruturais, mas com a ideia de que um dia se iriam autonomizar.

Até que ponto a Comissão Europeia tem interesse em deixar de financiar os cursos profissionais? Porque apesar de tudo está a financiar quase metade da oferta...

Eles não querem saber são os cursos profissionais ou outra coisa qualquer. Eles querem saber é das suas metas. E a verdade é que nós ajudamos a média europeia, eles não querem saber se é através dos cursos profissionais, dos psicólogos ou de outra medida. Nós é que temos de justificar que é aquela medida que lhes vai permitir ter aquele bombom. a meta é o fundamental. É mesmo KPIS a um nível macro, estamos a falar de gestão. Se chegas lá porque usas esta ou aquela medida, eles querem é ver resultados, que aquilo que vão financiar vai ajudar a chegar lá.

Qual é o papel da segurança social no programa dos CLDS?

A segurança social é a entidade gestora responsável pela medida de política social CLDS a nível nacional. Sendo que é uma medida de política que se refinanciou no âmbito do POPH do QREN e que se refina no âmbito do PT2020. É uma medida de política que começou em 2007, que tem vindo a ter várias gerações e várias alterações, mas é uma medida política.

Qual é o papel que assume também, no decorrer de um processo de avaliação?

Após 2 gerações dos CLDS, que são projetos de desenvolvimento comunitário e não se conseguem avaliar logo que acabem, foi pedido junto do POAT, em 2014, a ver se tínhamos uma avaliação de impacto da medida de política. Mas não se conseguiu concretizar devido a mudanças de tutelas e de governos. Entretanto, estavam a decorrer candidaturas no âmbito do POAT, que agora não tem muitas candidaturas para organismos fazerem estudos, mas no tempo do QREN tinha. (...).

Como isto já estava em carteira (...), foi integrado na lista de estudos, a ser promovida pela própria ADC, o estudo de impacto de uma medida, já tinha havido outras medidas e outros estudos destas abordagens, nomeadamente um estudo pelas abordagens territoriais, etc, etc. E, portanto, houve aqui uma avaliação da medida de política que vinha a ser transversal já há vários quadros. E ela é financiada no âmbito do PT2020.

Nesse âmbito, o caderno de encargos, como era uma área setorial do POISE, ele é desenvolvido pelo POISE, onde as avaliações são tripartidas entre todos os POs, o plano de avaliação é comum. Depois, onde é que se aloca financeiramente, é uma partição entre as várias assistências técnicas desses POs, há uma decisão que esse procedimento é lançado pelo POISE e foi pedido o contributo ao ISS para a elaboração do caderno de encargos e peças, mas com uma grande orientação técnica do ponto de vista do estudo e dos indicadores de estudo e várias outras questões próprias da avaliação – aquelas coisas todas, que eu não sou da área da avaliação. Mas é uma linha que a própria ADC segue. E com base nisso, nós demos alguns inputs ao caderno de encargos, e depois foram feitos alguns trabalhos de campo e alguma recolha de informação que não foi fácil por parte dos avaliadores, com algumas reuniões comigo.

Nesse contributo que o ISS ainda conseguiu dar para o caderno de encargos, recorda-se em concreto que contributo foi, ou em que sentido?

Não me lembro, não me lembro porque eu dou contributos a n cadernos de encargos.

E nessa construção desse caderno de encargos, pelo menos recorda-se de outras entidades terem também participado na própria construção do caderno?

Não sei, acho que foi o POISE que era anterior gestor, também o POPH, portanto, a autoridade de gestão, e a ADC não me recordo de mais ninguém. Não sei se não deram também contributos as autoridades de gestão regionais de Lisboa e Vale do Tejo e Algarve.

Falando ainda um bocadinho do caderno de encargos. Aquilo que era que era solicitado, era de facto, uma teoria da mudança, não é? E essa parte, portanto, estava fechada, não é? A Doutora Cristina também deu conta disso. Essas questões já estavam, de alguma forma alinhavadas.

Eu penso sim, que a teoria da mudança e uma série de requisitos, eu acho que havia...

E daquilo que é a sua experiência e perspetiva de alguma forma, tendo em conta esta avaliação que foi feita dos CLDS, qual é que considera que terá sido o papel da teoria da mudança nesta avaliação?

Nem consigo perceber se quer que lhe diga. Não consigo muito bem perceber na teoria da mudança, o que é que eles pretendem...com a análise da teoria da mudança. Porque aquilo é tão complexo, do meu ponto de vista, que eu acho difícil.

Então é como se o ISS sentisse que não tem um ganho com a execução da teoria da mudança, com o desenho da teoria da mudança. Qual é o ganho do ISS?

Eu acho que esse relatório tinha de ser visto e lido e comentado, e eu acho que a única pessoa que comentou fui eu, quando foi feita a apresentação pública e, portanto, aquilo tem ali uma série de recomendações e decisões que são decisões não tão técnicas, pela lógica da teoria da mudança que estava a falar, mas são decisões de amplamente de mudança de teoria de política. E, portanto, eu acho que estas avaliações para terem algum impacto têm de ser, têm de ter um alinhamento político estratégico. Não sei se estou a ser clara?

O que eu acho é que ninguém ligou nenhuma a esta avaliação. Porque se for ver, não sei se viu, se não viu, se olhar bem para quem respondeu... Se bem se recorda aquilo, tinha ali entrevistas de stakeholders relevantes nas diferentes fases... Se verificar a maior parte das pessoas não tiveram disponíveis sequer para a entrevista. Os elementos estratégicos relevantes não tiveram disponíveis para entrevista. E a entidade que era responsável daquela política pública também não teve disponível, porque eu penso que a única entrevista que eles conseguiram entrevistar em funções fui eu. Não entrevistaram sequer a minha chefia acima. E acho que também não estiveram disponíveis para entrevistar nenhum membro do governo que era responsável por esta medida de política. Portanto, isto é tudo muito ténue, pode-se ter muitos estudos e muitos impactos, mas se não houver um envolvimento dos verdadeiros stakeholders no processo, ela é ténue do que é que vai ser a teoria da mudança.

Então, o relatório aparece numa determinada fase, no final do ano de 2019, 2020, vésperas do covid, não é? Portanto, fazemos a simulação em dezembro, e eu lembro-me de ter dito à minha chefe, sobre as recomendações que ali estavam colocadas, que era anterior vogal do Conselho Diretivo, de dizer que o ISS, perante aquele documento, perante as suas estruturas, que são vários níveis, porque nós temos centros distritais, temos territórios, temos diretores de unidade, desenvolvimento social, diretores distritais, que era necessário fazer uma reflexão sobre este assunto. Portanto, eu acho que o tempo também não ajudou, porque nós fomos invadidos por uma pandemia e passamos a correr atrás do prejuízo, tentar pagar layoffs e resolver problemas. E, portanto, ninguém tratou deste assunto, do meu ponto de vista.

Houve ali algumas recomendações que, antes do relatório ter sido publicado – porque houve um relatório preliminar e na altura estava-se a desenhar o 4G, a Quarta Geração dos CLDS – que nós conseguimos fazer ali alguns ajustamentos, logo. Mas do ponto de vista técnico, ou seja, conseguiu-se aproveitar algumas daquelas recomendações para serem operacionalizadas. Uma verdadeira teoria da mudança, ela só é eficaz se for convertida politicamente. Eu acho que ela é importante dizer o que

é que é. Agora também tenho aqui uma opinião, que eu conheci bem os avaliadores do processo. Eu acho que os avaliadores também já vinham para uma predisposição e com um preconceito sobre o processo.

Sobre o próprio processo de construção da teoria da mudança?

Claro. Claro, porque as próprias pessoas que ganharam este caderno de encargos, foram pessoas que, formalmente ou informalmente, chegaram a ser consultados pelos então governantes naquele momento. Portanto, tinham ligações privilegiadas neste âmbito. E, portanto, tinham um próprio desenho e um entendimento do que é que isto devia ser. E o que é hoje o entendimento sobre os contratos locais de desenvolvimento social, não é o que se perspetivou que devia ser há 10 anos atrás, no seu início. Portanto, os CLDS têm vindo a ser enxertados, por diversos fatores. Porque um CLDS é uma coisa que se faz e que tem um embrião na rede social, por isso é que eu estou a dizer porque a Rede Social nasce com o Paulo Pedroso, com o Luís Capucha, portanto, eu fui entrevistada pelos 2, sei o que é que estou a dizer. E, portanto, eles têm predefinições do que é que deveria ser e alinham para aquilo.

E o que é a realidade?

Eu nem consigo perceber se a realidade é aquilo ou não. A realidade política não é aquela. Porque o problema é os fundos comunitários pagarem medidas de política, é que nós temos que enxertar em fundos comunitários coisas que não são exequíveis em 36 meses. Não têm uma flexibilidade de gestão que não se conseguem ter, porque depois nós temos a regras de gestão e procedimentos administrativos, que é uma carga brutal. Este é o problema, porque temos muita contratação pública, temos muito não sei quê...

É melhor do que não ter nada. Mas um contrato local de desenvolvimento social devia ser um chapéu transversal e muitas das vezes, eu disse isto já muitas vezes, ele até se choca com o que é um instrumento de DLDC (?). Porque não há contratos de desenvolvimento social sem haver máquina produtiva ou setorial ou económica, não é? Porque não se faz só desenvolvimento económico e social, trabalhando só o social. Podemos ter um eixo de Empregabilidade, mas se não tivermos outros fundos que possam ajudar, ou que podemos pagar a criação de um próprio emprego, ou uma coisa de pescas, ou uma coisa da agricultura, com uma responsabilidade transversal...

Agora eu acho que a teoria da mudança, que é isso que está a falar do ponto de vista do estudo, que faz sentido que nestes modelos de estudo. Mas eles têm de ser vistos - estou a falar disto como estou a falar, por exemplo, das inovações...

Uma questão: A teoria da mudança é geralmente feita pela equipa dos avaliadores, porque para se avaliar os impactos, é necessário haver um racional teórico que diga de onde é que vêm os problemas, quais são os mecanismos de mudança que tem de ser acionados e que aquelas medidas em concreto que estão a ser financiadas promovem para obter os resultados e os impactos. Esse racional não deveria ser construído pelos criadores da medida política? Não era pelos avaliadores...

Do meu ponto de vista, sim. Pronto, então é isso, por isso é que eu estou a dizer, ele tinha que ser feito, ou pelo menos pré-avaliado. Até podia vir com uma proposta, não é? Mas para se avaliar aquilo, ele devia de vir pré-avaliado, ou devia vir uma encomenda, quem encomenda o estudo dizer o que quer da teoria da mudança com essa lógica que está a falar. Mas não é feita, porquê? Porque fazem isso no estudo e queremos se teorias da mudança, que a Comissão Europeia quer ter impactos e obriga a este item, mas depois não há um envolvimento de quem define a política. Porque muitas vezes o que acontece é que acabam por pedir pessoa técnicas para o fazer, mas depois quem está em cima, não avalia.

A minha questão é que os avaliadores, geralmente, têm que avançar com essa parte, porque ela não existe...

Pois, já percebi

Ao existir, os avaliadores já não teriam... e se calhar é isso que está a falar, têm ideias preconcebidas. Eu, nos projetos em que estou envolvida, nós montamos a teoria da mudança com base em literatura científica anterior. E a literatura aponta num sentido. Pelo menos é assim que eu trabalho. Não monto uma teoria da mudança a pensar naquilo que eu acho que devia acontecer, mas naquilo que a literatura me diz que é aquilo que deve acontecer. Mas só faço isso porque o próprio programa operacional ou medida política não tem esse racional, porque se tivesse, eu só ia para o terreno verificar se aquela teoria da mudança se estava a verificar.

Não, não tem, não tem, não tem, não tem, não tem porque essa teoria da mudança devia existir quando eu concebo uma medida de política, ela devia estar explícita desde logo. Qualquer política pública devia dizer qual é que era o seu objetivo ao criá-la. E eu acho que muitas delas são meros despachos administrativos e não são efetivas mudanças, não são políticas públicas que se queiram gerar mudança.

Vou dar um exemplo: o PRR. Se nós lermos o regulamento comunitário, o PRR é um programa de recuperação e resiliência, face a um diagnóstico estratégico que as políticas têm de se monitorizar e tem que haver mudanças. Nos que eu acompanho, são investimentos puro e duro administrativos, como, por exemplo, comprar carros elétricos.

Portanto, aquela medida de política tem 4 investimentos que não têm um fio condutor. Por exemplo, tem lá uma medida que chama o radar social, que é idêntico ao do CLDS, que eu estou a ver se “desembrulho”, porque eu se calhar, vou ter nos territórios 2 processos em simultâneo a fazer a mesma coisa, um com financiamento do PT 2030, outro com o PRR. E estou eu aqui, a tentar numa lógica de política pública, arranjar um referencial que diga que eles são intervenções complementares. Mas ninguém me disse, nestas políticas, qual é a teoria da mudança, quais os objetivos e de que forma transformam os territórios onde estão inseridos. Esta é a questão que temos aqui. Porque depois o que eu tenho descrito como estratégia, ainda no caso do PRR, por exemplo, o que tenho é nada mais do que dizerem que eu vou ter de fazer um mapeamento dos idosos. Que teoria da mudança é esta?

A georreferenciação, não é um mapeamento, chama georreferenciação, das pessoas idosas ou com dependências. Foi um problema que se identificou, que não se sabia que durante o covid onde é que estavam as pessoas e quem eram as pessoas. Tudo bem. Mas para quê?

Pois isso não existe, isso prejudica-vos a vocês que implementam as medidas e depois prejudica os avaliadores...

Quem está em cima é que tem que definir. E se for ler, todas as portarias de política pública estão todas enviesadas, quer dizer, podemos ter algumas na área das economias e das empresas, aí conseguimos perceber, é para meter mais tecnologia, aumentar a produção, esses têm algum impacto, não é? Agora o resto é tudo muito ténue.

Voltando um bocadinho atrás, já falámos sobre isto, mas só para que, pelo menos para mim, que fique um pouco mais claro. Portanto estamos numa situação em que a teoria da mudança não é feita previamente. Não há este racional explícito, não é? No entanto, ela é exigida nas avaliações e os avaliadores têm de fazê-la, seja por que via for. Nesta situação, em que tem de ser feita para cumprir com o caderno de encargos, e daquilo que é a sua experiência, qual é que ainda assim é a vantagem ou benefício que esta que esta abordagem traz às avaliações, se é que algum, naturalmente?

Pois, não sei. Eu acho que teria sempre de haver uma perspetiva de que eu quero saber, pode não ser a teoria da mudança, mas devia haver pelo menos uma coisa, que é perceber o que acontecia em determinado momento, antes da intervenção política pública e o que é acontece a seguir. Não sei se isso tecnicamente se diz teoria.

Estou aqui a pensar no CLDS, num determinado território com determinadas características, que era vulnerável. Vai-se fazer ali uma intervenção. Faz uma intervenção de x anos, 3, 6 ano. E depois passados 6, 7 anos quer-se perceber: melhorou, não melhorou, piorou, mas também pioraram as condições económicas e se não tivesse havido intervenção, ainda tinha sido pior. É isto, um bocadinho, que nós queremos medir no impacto. Isto é o que eu acho que é mais razoável. Imagine que não melhorou, não foi eficaz, gastou-se o dinheiro, mas nada aconteceu. Então eu tenho que perceber o que é que estrangulou ou não a existência disto. E isto tem que gerar contributos, porque os pressupostos daquele programa eram estes, e agora deveriam ser reajustados? Se isto é um modelo de teoria da mudança ou não, não sei, já estou aqui baralhada. Vocês

tecnicamente saberão dizer melhor do que eu, mas isto é o que eu acho que era necessário para qualquer político tomar decisões.

De alguma forma, também daquilo que é o meu conhecimento, a teoria da mudança deveria dar isto... Procuramos ter com as pessoas com quem vamos entrevistando, precisamente, esta reflexão: o que é que tem sido feito, como, porquê, que vantagens e desvantagens é que tem trazido?

Não trouxe nada, ficou na gaveta. É isso que objetivamente que eu lhe posso dizer, ficou na gaveta, mas eu acho que, enquanto não existir outra, eu ainda não perdi a esperança de olhar para aquele exercício e ver o que é que ele diz. Porque, só para lhe dar um exemplo, eu parei o modelo de acompanhamento dos CLDS, porque o modelo que lá estava descrito tinha vicissitudes, não me parecia que fosse correto.

Quando diz lá, no relatório da avaliação e nas recomendações?

Sim. Eu parei, porque estar a fazer uma coisa só por fazer, que não tem impacto, que tem de ser repensada. E, portanto, assumi e disse que não vou estar a dar orientações a ninguém para fazer isto, porque isto precisa de se parar, pensar, olhar, ler o que é que outros viram. Portanto, eu ainda não perdi a esperança de olhar sobre este processo, para que num próximo - já não no 4G, mas no próximo 5G – se possa ter aqui algum apport.

Portanto, eu acho que ele é um instrumento que, apesar de estar datado em determinado momento, ele continua a ser útil porque eu não implementei tudo. Eu não implementei aqui, mas tenho que implementar mais à frente, nomeadamente no PT 2030.

Agora, eu também gostava que alguém tivesse olhado para aquilo quando estava a desenhar o PT 2030. E ninguém olhou, politicamente.

E nesse sentido, portanto, quando foi o desenho do PT2030, o papel do ISS...

O papel do ISS é o quê? É dizerem que têm os CLDS, que têm um dinheiro idêntico ao anterior e, portanto, os CLDS continuam a ser a mesma coisa, não podem ser uma coisa diferente, que é o que lá está escrito. O que está lá escrito, que eu leio, é se se querem ter determinadas ações avulsas durante 3 anos, ou se querem ter uma coisa mais estruturada, que não seja só naqueles eixos. Quando estão a desenhar um programa têm de desenhar medidas de política, e dizer que umas entram, umas saem, outras continuam. Aqui o que me disseram foi que há uma continuidade.

(...)

Vamos ter uma coexistência territorial, dos CLDS e do Radar, com as mesmas entidades a serem responsáveis pelas políticas, que vão ser as autarquias a intervirem em desenvolvimento social, onde num trabalho os idosos e as dependências e outro, nos CLDS, onde tenho o eixo 3, para o envelhecimento. Isto não me faz sentido. Isto tem de ser devidamente adaptado. O mesmo no eixo, 2 (...). [fala sobre PT2030, etc, etc]. Não há uma estratégia.

No caso dos CLDS, quando começaram, houve um diagnóstico prévio à sua criação?

Houve. Há um diagnóstico territorial, feito pela Rede Social. Esse é pontapé de saída para os CLDS.

Fala sobre o PROGRIDE e sobre o contexto em que surgem os CLDS, com vários projetos a trabalhar mtas populações que se cruzavam; isto dispersava os recursos e viram que o PROGRIDE fragmentava a intervenção. Fazia sentido ter um chapéu de intervenção agrupado em eixos: emprego; infância; idosos; informática. Concentra-se tudo num único processo e com base no diagnóstico da Rede Social. Tinha uma lógica de concentração de recursos num mesmo território, com diagnóstico territorial. Mas viu-se que territórios com mais qualificações e mais técnicos tinham sempre intervenções, enquanto territórios mais carenciados, nunca tinham intervenção. Então interveio-se nestes. Foi o princípio do diagnóstico, que permitiu definir os territórios de intervenção. Isto permitiu intervenções distintas nos territórios. (...)

O contrato de desenvolvimento social deve ser feito com base no diagnóstico, o PDS (plano de desenvolvimento social). E do que não temos financiamento ir buscar aqui. O que quer dizer que os CLDS dizem uma coisa que ainda ninguém percebeu, nomeadamente, que têm dimensões obrigatórias. Mas se eu tiver dimensões obrigatórias, e essas já estiverem a ser financiadas por outras linhas de financiamento, eu posso não utilizar o financiamento e utilizar o financiamento para outra coisa, para tentar majorar. Mas os senhores presidentes de câmara não são muito inteligentes, porque se fossem, usavam isto e não usam. (...).

Há também dificuldades em trabalhar em parceria, há uma lógica de trabalho em rede que não é conseguida. Em teoria é magnífico. Mas na realidade não funciona. Os projetos podem ter parceiros, 3 ou 4 no máximo, nós tivemos de reduzir. Porque era ingerível. Eles não se entendem. Trabalhar em parceria é eu deixar de pensar só na minha instituição e pensar no bem comum do território, e isto não existe. Isto é filosofia barata. Portanto, operacionalmente, territórios que tinham 11 parceiros andavam à batatada.

A maior parte dos projetos que temos em intervenção são monointervenções. Não têm parceiro nenhum, porque trabalhar em parceria é muito difícil.

Isso desvirtua aquilo que era o objetivo do programa?

Pode ou não desvirtuar. (dá outro exemplo, que não os CLDS e fala sobre isso, em como se uniram num dado território e estão a servir a população; critica a forma como as pessoas gerem os projetos e alocam mal o financiamento).

Quando foi criado o programa, estava previsto algum sistema de monitorização?

Esse é que é o problema. Inicialmente, quando o programa foi pensado, foi criada uma enorme bateria de indicadores. Havia uma prestação de serviços com o ISCTE, para desenvolverem isto, há relatórios, mas eu na altura não tinha funções nesta área. A verdade é que isso caiu tudo. E quando eles vêm fazer a avaliação, eu só tenho papel. Não havia um sistema de gestão de informação. O sistema de informação que existia no POPH era um sistema de gestão de informação em que eu carregava indicadores muito sucintos, como, por exemplo, número de pessoas, por eixo, sem detalhe para quem quer fazer outros estudos. E não há uma bateria de indicadores brutal, nem há no PT 2020. Portanto, medir estas medidas de política, obrigava a que o próprio plano de ação tivesse uma série de indicadores, metas, etc, e fosse altamente modificado e fosse feito em função de resultados, que era também esse o principal objetivo: era dizer que eu tinha de fixar para cada objetivo uma meta. E agora não há. O plano não está assim desenhado. Há atividades e tem pessoas a atingir. Não tenho metas que eu quero alcançar.

Neste momento, se eu for um território que só sou vulnerável no emprego, eu só tenho de fazer o eixo do desemprego. Não sou obrigado a ter ações para a área das crianças ou dos idosos. Porque não tenho esse tipo de vulnerabilidade.

No início, o financiamento era o mesmo, entretanto, mudou. Mas só mudou no 4G. Ou seja, o financiamento era constante. Nós dávamos 450.000 EUR. E ISTO foi uma coisa que o avaliador alertou, porque eu dava o mesmo financiamento a Barrancos que dava à Amadora, ou a Gaia. Neste momento, não. Neste momento, o financiamento é em função do número de vulnerabilidades que o território tem e da população que lá existe. Ou seja, é um fator combinado. E foi majorado para as áreas do interior. Esta já foi uma alteração que fizemos com base numa recomendação da avaliação.

O que é que explica a não execução ou não prossecução do sistema de monitorização de informação?

Primeiro, porque nós não temos sistemas de informação e nós dependemos do instituto de informática para ter sistemas de informação. Depois, porque quando se montam as coisas, montam-se as coisas antes de existirem instrumentos para as montar. Dou-lhe um exemplo, exemplo, na próxima semana, vão fazer sair um aviso do PRR, que não há sistema de informação para entrar candidaturas. Quem está na política, quer fazer sair a medida de política com calendários que não são ajustados. E depois não se consegue correr atrás do prejuízo. Nós púnhamos tudo em folhas de Excel durante anos. Não pomos agora, porque existe uma plataforma do PT 2030. Mas não é uma plataforma da política pública. Só recolhe indicadores para o PO e de reporte à Comissão. Se formos olhar para a política pública, a política pública o que diz é que quem tem responsabilidade por verificar e ter a monitorização do plano de ação e de toda a execução do CLDS e o próprio Conselho. Portanto, não existem um modelo nem uma bateria de indicadores que eu possa ter. E este é um problema para a avaliação, é que cada autarquia tem indicadores próprios. O próprio diagnóstico, o próprio PDS e a própria metodologia divergem de território para território, portanto, eu criar um instrumento que seja transversal é ruidoso, ainda que diga respeito às ações obrigatórias. Quanto às ações obrigatórias, a recolha de dados, quantos destinatários e quantos é que atingi, isso eu recolho. Mas ninguém quer ter impacto com a medida, nem o próprio PT2030 e PT2020. Porque se formos ver qual era a medida de impacto que tínhamos negociado, fomos negociar e passar a outra e dizer que temos atividade concretizadas, garantindo 70% dos destinatários, é o que diz o indicador. O que o indicador inicialmente dizia é que tinha de ter 5% da população ativa envolvida numa medida de emprego ou formação, isso é que é o impacto, mas não a quiseram meter, já saltou fora. Acho que ninguém quer impactos. Lá em cima, ninguém quer, desculpe lá.

Qual a expectativa do ISS quanto à avaliação?

Acho que a avaliação dá os dados ao ISS que ele precisa para corrigir isto, do ponto de vista operacional; e também dá acima. Ninguém quis foi tratar disso, nem no ISS, nem acima. Ele tem o pontapé de saída. O Gabinete pediu-me a avaliação dos 4G e eu mandei o estudo; não queriam impacto?

Sobre AIC...

Não foi feito por vários fatores, pq precisavam dos nomes das pessoas alvo da intervenção e eu disse que essa info constaria nos projetos das pps entidades, nos dossiers tecnicopedagogicos. Depois colocou-se uma questão de proteção de dados e acordaram com ADC e POISE de não fazer. Esta parte dos dados também tem de ser garantida previamente, quando as pessoas participam e assinam um termo.

A gestão da avaliação foi mt mais feita entre a equipa da avaliação e ADC do que com o ISS, se estou a perceber...

Sim, a monitorização da avaliação foi feita com o POISE, com o Paulo Rafael e outra técnica. Ele foi o chefe da avaliação. A técnica não me lembro do nome. Mas também o Domingos; o Paulo Rafael depende do Domingos.

Se tivesse oportunidade de repetir a avaliação e se pudesse, o que alterava? Fazendo um balanço, o que alterava?

Eu simplificava. Acho que a quantidade de questões naquele caderno de encargos... acho de uma densidade brutal. Todos aqueles itens... Até acho que há dimensões que se repetem. Era preciso simplificar, para todos os intervenientes. Agora percebo que aquilo tem uma metodologia, uma matriz teórica. Não sei se é possível simplificar...

Acho que a quantidade de perguntas, os relatórios, é tudo muito denso. Acho que este modelo de exercícios de avaliação são mt mais densos e de difícil leitura – não nas conclusões e recomendações – que eram outros noutros períodos de programação. Acho que há uma mudança qualquer, não sei se é por exigência comunitária, porque ele tem uma estrutura que é comum, portanto, eu acho que é obrigatória. Acho que a ADC impõe aquilo a toda a gente. Acho que não é o meu Core Business, mas para quem lê, aquilo é difícil. Lembro-me de outros na altura do QREN, que tinha perguntas chave e respostas chave, era de muito mais fácil leitura. Agora temos 10 dimensões, em cada dimensão não sei quantas perguntas, aquilo é de uma complexidade... Matrizes atrás de matrizes. Deve ser brutal para os avaliadores, para a equipa. Uma carga administrativa

esquizofrénica, que eu acho que sumida, acho que a maior parte das pessoas não tem capacidade para ler. Acabamos por ler o sumário executivo.

Relatório/transcrição da entrevista a responsáveis na AD&C

Qual o papel de ADC nos processos avaliativos de programas financiados pelos Fundos europeus?

P1: A nossa função está inscrita no modelo de governação dos Fundos; e é coordenar o sistema de avaliação e monitorização dos fundos. Em concreto, coordenamos a rede de avaliação e monitorização do PT2020, porque tudo se decide nessa rede. Está presente ADC, todos os POs, mais as entidades de agricultura e pescas. Nessa rede desenvolvemos o plano global de avaliação, que orienta todo o trabalho de avaliação durante o período de programação, estamos presentes no plano. Em concreto, na implementação desse plano, somos responsáveis por algumas avaliações, coordenamos algumas, outras são coordenadas pelas próprias autoridades de gestão. Relativamente às que coordenamos, somos nós que coordenamos o grupo de acompanhamento, que lançamos os concursos e que gerimos todo o contrato. Relativamente às avaliações que são feitas pelas autoridades de gestão, nós acompanhamos o processo como parceiro ativo, fazemos parte dos júris de concurso (...), é uma relação mais de acordo com as necessidades que as autoridades de gestão vão tendo, e somos parceiros no grupo de acompanhamento.

Foi dentro deste PGA que surgem as fichas de avaliação? Esta questão das fichas surgiu noutra entrevista e disseram-nos que as fichas fechavam, no fundo, as avaliações. podem falar-nos disso, como são construídas, objetivo geral...

P1: Nós consideramos as fichas uma peça fundamental porque é o que nos permite antecipar as avaliações. As avaliações, sobretudo as contrafactuais, precisam de um trabalho prévio, de se conseguir mobilizar outras entidades relativamente aos sistemas de informação. Quanto mais cedo fizermos este trabalho, melhor. Um exemplo: um contrafactual em 2016/17, e demorámos um ano a mobilizar todas as bases de dados que queríamos utilizar... levámos um ano a fazer um protocolo com o INE. Tudo antes de lançarmos a avaliação. As fichas permitem, por um lado, ter um trabalho de planeamento a 7 anos, portanto, conseguimos identificar logo quais são as avaliações e o que queremos saber, as metodologias a usar e todo o trabalho preparatório necessário. As fichas, cada uma delas pode ter histórias diferentes; pode ser ADC que propõe, outras as entidades de gestão... No plano, as fichas é o que demora mais a fazer.

P1: Há uma coisa importante para perceber: na Comissão Europeia fazem uma monitorização do plano anualmente; mas às vezes as avaliações atrasam-se, porque as próprias intervenções objeto de avaliação também se atrasam. Então só se pode avaliar quando há resultados para serem avaliados. E, por outro lado, no PGA, que é um plano a 7 anos, o plano é revisto anualmente; e essa revisão é mesmo importante – nós não temos a expectativa que o plano se mantenha igual, o plano é flexível e responde às necessidades que vão surgindo. (...). Isto significa que podem surgir novas avaliações inicialmente não previstas, ou podem sair outras avaliações, cujas áreas já foram cobertas por outras avaliações.

A Comissão Europeia, em termos do desenho das avaliações, tem algum contributo, faz alguma recomendação?

P1: Não, não.

A Comissão Europeia não tem, então, nenhuma especificação em termos de abordagem metodológica aos objetos de avaliação?

P1: Diria que há 3 formas da CE interagir com os Estados Membros. Uma, é através dos Regulamentos, e para o PT2020 tínhamos de fazer Avaliação de Impacto e analisar a eficácia, eficiência e impacto das intervenções. E que tínhamos de avaliar o impacto de todos os objetivos específicos. (...). Por exemplo, o abandono escolar é um objetivo específico; o aumento de diplomados do ensino superior é outro objetivo específico. Mas o Regulamento não era claro e isto veio a atenuar-se. Ou seja, todos os objetivos específicos tinham de ser avaliados, mas não tinha de ser avaliação de impacto, podia ser avaliação de processo, por exemplo. A segunda interação é através de Guias de Orientações. Aqui já entra a questão das metodologias. E aqui as avaliações de impacto já têm de ser ou AIC ou ABT, remetendo muito para o EVALSED, o guia onde estão explicitas as várias

metodologias. E depois, um terceiro nível, através das Redes de Avaliação e Monitorização, onde há cerca de 2 reuniões por ano, do FSE e do FEDER, onde esses temas vão sendo discutidos e apresentados.

Em que momento sentiram necessidade de formalizar as fichas do PGA? E qual a influencia destes guias da CE, até que ponto se basearam nos mesmos ou noutras experiências?

P1: Em relação às orientações, elas não são obrigatórias. Mas não há nada nas orientações que não se aplique ao caso português. Não é que aquilo seja considerado obrigatório, mas os Guias saem de estudos, são fundamentados noutros estudos. (...). Os Guias dão um caminho, que nós adaptamos.

P2: As orientações da CE deram um contributo para a construção do nosso Guia de Orientações, associado ao PGA, que tem um conjunto de linhas gerais sobre cada uma das abordagens, portanto, ajudou a estruturar. E ajuda as próprias Autoridades de Gestão a desenhar os próprios cadernos de encargos e a estar a par das metodologias que, supostamente, os avaliadores farão recurso. Este nosso Guia é mais genérico, sobre AIC e ABT.

P1: O nosso Guia foi criado, sobretudo, para ajudar as Autoridades de Gestão a desenhar os próprios cadernos de encargos.

P1: Mas relativamente às Fichas, elas pretendem ser o resultado do trabalho que se faz na Rede. São observadas as necessidades de avaliação, que dão origem a cada uma das avaliações que são definidas. Depois, dentro dessa avaliação há um conjunto de decisões que têm de ser tomadas: qual o âmbito e o objeto de avaliação... as Fichas surgem para definir e sistematizar as avaliações que vão ser desenvolvidas. O Plano e as Fichas são vistos como um trabalho evolutivo. Há uma primeira versão e depois as fichas têm de ser atualizadas 1 ano antes de ser lançada a avaliação. No início as Fichas são mais abertas e vão sendo consolidadas ao longo do processo, que termina com o caderno de encargos. E é um documento aberto para o mercado. E o documento da avaliação é um documento obrigatório. Há um elevado grau de liberdade para as fazer. Foi opção de PT fazer um Plano Global, porque já tinha esse histórico, da altura do QREN, antes de ser obrigatório. Nós consideramos as Fichas um elemento muito relevante, porque através delas identificamos as necessidades de informação, que temos de preparar com muita antecedência.

P2: As Fichas até são uma base para negociar, porque por vezes, várias Entidades têm interesse num tipo de avaliação que depois podem ser coordenadas. São uma base de reflexão, de ver confluências de interesses, possibilidades de combinar recursos que podem ser do interesse de vários organismos.

P1: Se repararem, a maioria das avaliações são avaliações transversais, que incluem mais do que 1 programa. Se isto não fosse articulado e combinado numa Rede, seria impraticável.

De que forma são equacionadas e decididas as abordagens metodológicas?

P1: Importa, talvez, rever aqui os conceitos. Nós definimos na nossa equipa que método é Método Contrafactual e Método de ABT; abordagens, é cada uma das metodologias dentro desses dois chapéus. E então as técnicas de recolha da informação... Nós não definimos as abordagens. No PGA só definimos se tinha ABT ou AIC. Na realidade, não foi muito difícil. Tudo o que conseguimos perceber, há 7 anos, que era possível fazer com AIC, foi feito com AIC. O resto com ABT. Isto de uma forma genérica. Não definimos nada no que respeita às técnicas; ficou sempre em aberto para ser proposto pelos candidatos nas propostas.

Há, então, negociação sobre as abordagens e técnicas?

P1: Na verdade, não há bem negociação, porque as equipas propõem e há uma avaliação das propostas. Com a equipa vencedora, depois não negociamos. Podemos, de alguma forma, solicitar na reunião de arranque alguma pequena alteração, pode haver uma reflexão em conjunto, pedir uma entrevista aqui ou substituir um inquérito por outra técnica. Ou seja, é aceite a proposta vencedora, porque foi essa a vencedora. Em concurso público não há negociação, ganha a proposta vencedora.

Que critérios são usados para escolher um método ou outro?

P1: Na altura em que se fez o Plano, com o tempo e capacidade que se tinha, definimos que o que se sabia no momento, que tínhamos condições para fazer AIC – temos de garantir que é possível fazer, que há bases de dados, grupos de controlo. Quando conseguimos garantir isso, então definimos AIC. As restantes ABT. Não quer dizer que em certas avaliações não se tenham pedido ambos os métodos, como na Avaliação do Abandono Escolar.

Sobre a ausência de AIC na avaliação dos CLDS?

P2: No caso dos CLDS não estava previsto AIC. Mas a equipa propôs que fosse feito e ganhou, porque tinha mais qualidade, coerência, etc. Fazer AIC neste caso, tínhamos muitas dúvidas. Havia dificuldade em definir unidades de observação, problemas com informação quantitativa, e, no fundo, dificuldade em sustentar que dado efeito se devesse à intervenção dos CLDS. (...).

P1: Isto vem ao encontro do que estávamos a dizer: se tivéssemos previsto AIC, teríamos feito o tal trabalho de casa. Por vezes, mobilizar a informação necessária demora tanto quanto a própria avaliação. Aqui não se previa, mas a equipa propôs e tentou-se, mas não deu.

O pedido de AIC implica que tenham uma ideia muito concreta sobre o objeto de avaliação e as condições de execução. Mas no caso de ABT, há múltiplas: pode ser TdM, pode ser Abordagem Realista, de Contribuição... Parece que há o AIC que é altamente especificado, mas depois muita liberdade nas ABT...

P1: Mas nós damos a mesma liberdade dentro da AIC, porque nós não definimos se é a Diferenças das Diferenças... dentro do Contrafactual, em alguns casos, podemos achar que há uma abordagem que faz mais sentido, mas não definimos.

Porquê que o AIC deve estar nas avaliações? Já pensaram, em alguns casos/programas, em especificar abordagens dentro das ABT?

P1: Sim, já pensámos e refletimos nisto. Mas volto a dizer que dentro da AIC não somos prescritivos relativamente à abordagem. Em nenhuma avaliação dissemos qual era a abordagem. Na realidade, dizemos o método, AIC ou ABT, mas nunca dizemos qual é a abordagem em cada um deles. A questão é: é verdade que no Contrafactual já sabemos as fontes de informação, sabemos que são fontes administrativas. Nunca pensámos em fazer inquéritos, fazemos com fontes administrativas. E aí fazemos esse TPC. Essa é a diferença entre AIC e ABT. Talvez outra diferença: realmente, em termos de AIC, temos boas práticas, temos muitos estudos nacionais e internacionais sobre AIC; temos muitos bons exemplos. Na ABT não temos a mesma experiência. (17'15)

P2: Mas se calhar convém clarificar uma questão: de facto, há 7 anos, com o que sabíamos, aquilo que tentámos foi que, sempre que houvesse informação, fazer Avaliações Contrafactuais. Agora, as Avaliações Contrafactuais não o método de excelência. Elas dão um tipo de resposta, mas temos a noção que se quisermos perceber para quem funciona, em que condições ou o que é que explica que tenha funcionado, temos de ter outro tipo de avaliação. Neste momento, já é claro que não é o nosso método de eleição. A questão aqui é que AIC tem exigências de preparação de informação que exigem muita antecipação, por isso é que talvez haja esta ideia de que nós preparamos muito bem, mas não é assim... não é um método privilegiado.

P1: para AIC temos de ser nós a dar informação, enquanto para ABT são as equipas que recolhem.

Sobre AIC: vamos estar sempre limitados às bases de dados e indicadores existentes, é isso? Ou pensam que é possível criar novos indicadores, por exemplo, a partir da monitorização?

P1: É uma questão muito pertinente. No PT não fizemos isso, porque não tivemos oportunidade de o fazer em tempo útil. Quando arrancámos com o Plano, os sistemas de informação já estavam a funcionar e tivemos noção que não dava. Neste momento estamos a tentar. Estamos a pensar nas avaliações que gostaríamos de fazer, em algumas, em que faz todo o sentido. Antes dos programas começaram, queremos identificar necessidades de informação e recolher essa informação em sede de candidatura. Isso era o ideal para nós. Vamos ver o que conseguimos fazer nesse sentido.

No caso do POCH, fez-se AIC com os cursos profissionais. Mas será que o AIC só dava para isso? Foi satisfatório? Daria para cobrir mais medidas?

P1: Os dois métodos, AIC e ABT, acho que se complementam. Não podemos pedir a um, o que não pode dar. ABT nunca vai dar números, não é possível.

O AIC tem uma ação localizada, mas as ABT, não dando um número, conseguem englobar o conjunto; o AIC não...

P1: na TdM, a ABT consegue percorrer toda a cadeia, enquanto o Contrafactual vai definir uma seta entre duas caixas. Podemos fazer vários Contrafactuais.

P1: Voltando à pergunta de há pouco, é importante situarmos as questões. Se fosse agora, se voltássemos agora, podíamos fazer AIC não só aos Cursos Profissionais, mas também aos Cursos de Aprendizagem, aos Psicólogos, à Formação de Formadores... Na altura tínhamos um estudo e um piloto em AIC, no QREN. Aquilo foi o que, na altura, achámos que era possível. Não queríamos propor nada que não achássemos que fosse possível. Tínhamos segurança e informação para fazer. Mas queremos crescer e estamos a lançar estudos nesse sentido...

No caso do POCH, são medidas financiadas pelo Programa Operacional, mas a gestão das medidas é feita pelo Ministério da Educação, ou do Trabalho, etc... isto por causa dos dados; na Educação, a responsabilidade da recolha é do Ministério da Educação se quiserem que sejam recolhidos outros dados, têm de articular com o Ministério da Educação, certo? E outra questão, a TdM tem um princípio muito de aprendizagem com a prática, para a perceber a lógica da intervenção e depois os stakeholders que estiveram desde a conceção até à concretização dos mecanismos de mudança se o que tinham previsto foi feito e se resultou. Depois como é que se dá a aprendizagem das instituições, dos stakeholders, face aos resultados da avaliação? Neste caso, por exemplo, do Ministério a dar dados, e do outro lado a aprender com a TdM

P1: Sobre o Contrafactual, faz todo o sentido conseguir essa articulação com os gestores, com os detentores das bases de dados. No PT2020, o que tentámos fazer foi: isto é o que existe; tínhamos de fazer com o que existia. Tivemos uma avaliação em que conseguimos usar 5 bases de dados, foram as Contas Nacionais, as Patentes, os Quadros de Pessoal, os investimentos em PME, era 5 bases de dados. Ainda deu origem a papers, porque a originalidade daquele Contrafactual foi ter juntado aquelas bases de dados, e não foi uma preocupação nossa. Enquanto a questão das nossas bases de dados, dos Fundos, nós conseguimos incluir informação de candidatura, sim, essa pensámos. Mas ainda estávamos numa fase inicial. Neste momento estamos mais preocupados em incluir nos nossos sistemas, dos Fundos, porque achamos que podemos fazer mais.

P1: Quanto à outra questão, diria que aí é muito diversificado. No acesso aos dados, sentimos muita abertura. No início mais, depois veio a questão da Proteção de Dados, que criou alguns constrangimentos, mas foi resolvida e os dados são anonimizados. A questão da aprendizagem depende muito das entidades. A DGEEC é um bom exemplo, até tem essa vocação.

Até que ponto, por exemplo, a TdM contribui e envolve realmente os stakeholders? Será que interagem da forma certa, mais adequada?

P1: temos de dividir em várias fases. Na fase de ceder informação são muito ativos. Depois temos a questão do envolvimento – há fases que estão mais bem conseguidas que outras. Há uma fase que pode ser mais desenvolvida, que tem que ver com o envolvimento no próprio desenvolvimento do caderno de encargos. Nós constituímos o Grupo de Acompanhamento no momento em que enviamos o primeiro draft do caderno de encargos. Mas não fazemos nenhuma reunião; enviamos o caderno de encargos e solicitamos comentários. Achamos que ao não fazer reunião, não estamos a envolver devidamente. É o balanço que fazemos a posteriori. Podemos envolvê-los mais logo desde início. Depois temos o Grupo de Acompanhamento que funciona e reúne aquando de cada um dos Relatórios – inicial, intermédio e final. A nossa experiência é que parece que vai havendo uma aprendizagem, os GA são locais de aprendizagem. Claramente onde as entidades participam mais é quando começam a existir resultados. Participam muito pouco, ou menos, no Relatório Inicial, que é metodológico; e mais no Relatório Intermédio, quando já há resultados. (...). Depois temos a questão do follow-up. Esta é uma área que começámos agora com

o PT2020, fazemos um documento de follow-up que enviamos para as Entidades responsáveis, com cada uma das recomendações e solicitamos que nos façam uma apreciação às recomendações. Ou seja, se a recomendação faz sentido, se vai ser implementada e se sim, em que moldes. Consideramos que ainda há muito trabalho a fazer aqui, com maior envolvimento das entidades. Também envolvimento na questão da TdM. Nós, de algum modo, delegámos nas equipas de avaliação esse envolvimento. Pedimos no caderno de encargos que as equipas desenvolvessem a TdM de forma consensualizada, ou seja, teriam de fazer workshops de construção da TdM. Não consideramos que tivéssemos casos de boas práticas nesta matéria. Achamos que não foi bem conseguida, esta parte.

Os recursos dedicados à TdM são suficientes? Ou seja, a TdM é apresentada na fase inicial, o que pressupõe revisão da literatura, um conjunto de entrevistas a atores-chave, envolvimento dos vários executores e, idealmente, de quem gere as medidas, desde gestores intermédios e finais, são feitos workshops, e como disse, consideram que não têm grande boas práticas, mas parece haver um tempo maior do que aqueles, normalmente, 3 meses iniciais em que se estão a montar as equipas...

P1: Não, claramente não. No mundo perfeito, nós teríamos uma TdM e todas as variáveis que eram necessárias para avaliar estavam no desenho do programa. Mas sim, consideramos que não foi fácil. No balanço que fazemos, acho que temos de alterar e já estamos a alterar. No próximo período vai ser completamente diferente. Vamos distinguir as fases. Ou seja, quando encomendarmos uma avaliação, já temos a TdM construída. Neste momento estamos a fazer estudos para construir a TdM antes.

Nós gostaríamos de fazer muito mais trabalho preparatório, com TdM, necessidades da avaliação, já com base na TdM, e identificação de informação necessária para responder aos métodos.

Quando vão começar?

P1: Já estamos a construir as TdM e a identificar necessidades avaliativas.

Essa construção da TdM encaixa nas Fichas de Avaliação, ou é um exercício à parte ou paralelo?

P1: Antecede. Na fase da ficha de avaliação, já é suposto n'ós termos as TdM e percebermos claramente as áreas que temos mais informação e quais os que devemos incidir, porque temos menos informação, e devemos incidir mais as avaliações.

P2: Isto não quer dizer que as TdM desenhadas não sejam sujeitas a revisão e que nós não desejemos ver, por parte dos avaliadores, um trabalho que inclusivamente passa por esse envolvimento. Uma das razões para a insatisfação não tem a ver só com o tempo, tem a ver com a maneira como se conduz o trabalho empírico e de recolha para construir a TdM. Na perspetiva de quem quer construir um Guia, uma das áreas de fragilidade nas equipas é o domínio de metodologias participativas que permitam construir as TdM; no próximo ciclo, alimentar/densificar as TdM. A forma como tem sido conduzido este trabalho parece-nos muito pouco indutor de contributos para a TdM. Esta é uma das áreas que consideramos frágil e gostaríamos de ver desenvolvida, ver o envolvimento dos parceiros na revisão e densificação da TdM. Não temos é o exercício de começar tudo do zero.

Estou a ver que isto pode trazer vantagens: já há um envolvimento inicial por parte dos stakeholders; por parte das equipas de avaliação talvez também já seja mais fácil, porque já está identificado um conjunto de atores; que outras vantagens e eventuais desvantagens? Por exemplo, poderá limitar ou constranger a equipa de avaliação?

P1: Já traz também a visão do próprio programador. Penso que não traz desvantagens. Se dissemos que estava ali uma TdM e tinha de ser essa, sim. Mas a nossa ideia não é essa. Sabemos que a TdM ainda é pior que os conceitos, há n formas de TdM. Usamos um modelo standard do quadro lógico, mais com mecanismos e pressupostos. Mas o que consideramos é que deve haver uma adaptação da TdM até à abordagem que vai ser usada. Por exemplo, para fazer uma Análise Contribuição é necessário um tipo de TdM; para fazer a Análise Realista é outro tipo, outro modelo de TdM... A ideia é que o quadro lógico

do programador quando montou a iniciativa/intervenção esteja lá. Mas vai ter sempre de ser adaptada à metodologia que se quer implementar.

P2: Tivemos dificuldades neste ciclo inclusivamente com conceitos base. Coisas tao simples como o que é um resultado e o que é um indicador. Às vezes as nossas interações com as equipas passavam por coisas tao básicas como o que é que se inscreve na TdM e o que é informação adicional. Então, o que resolvemos foi, pelo menos, estabilizar essa base e dar esse pontapé de saída. O que vier a seguir, vai enriquecer. Também se quis alguma padronização, obviamente sujeita a revisões. E também é um bom processo para nós, que nos metemos no lugar de quem é interpelado para fazer. Tem essas duas visões. Mas a partir do momento em que a base é estabelecida, consideramos que os avaliadores têm um papel em densificar e perceber quais são os mecanismos. Não só testar, mas ajustar em função de contactos posteriores com stakeholders ao longo da avaliação.

P1: Se somos pouco prescritivos, devíamos ser mais prescritivos, se somos mais prescritivos, devíamos ser menos prescritivos... Não é fácil. Compreendemos o lado dos avaliadores, mas não é fácil. Qualquer solução tem vantagens e desvantagens. Onde gostaríamos mesmo de melhor era nessa parte do envolvimento dos stakeholders na reconstrução e validação da TdM. Gostávamos mesmo de dar um salto nisto. Não é fácil. Por vezes são intervenções muito complexas. Quando são várias intervenções numa TdM, torna-se complexo. E fazer em Zoom, por exemplo, a reconstrução de TdM com stakeholders não é fácil, sabemos que não é fácil. Os stakeholders não têm que saber que estão a falar de um resultado ou de uma realização, ou de um indicador. Têm de saber é como é que funciona, o trabalho deles é esse, não é fazer quadros e esquemas, temos de ser nós e as equipas a chegar aí, a tirar partido do conhecimento que os atores têm. Quando lhes apresentamos um esquema, é muito difícil perceberem. Às vezes até nós que já estamos mais habituados, estamos horas a partir pedra em duas caixas: uma diz que não é aquela palavra, outra diz outra... Temos de arranjar uma forma de trazer esse conhecimento dos stakeholders para a construção da TdM, usando os conceitos teóricos dos modelos de TdM. Mas acho que temos de separar isto, o conhecimento de construir uma TdM, do conhecimento do funcionamento da medida.

Usando o POCH como referência, que tem várias medidas, quando visualizamos o modelo da TdM, vemos duas coisas: a teoria do programa, ou seja, a visão que o POCH tem de como atingir os objetivos e quais são as medidas que são incluídas, e depois, dentro das medidas, a teoria de como é que aquelas medidas chegavam aos mesmos objetivos. A questão é: será que o programa está a ser avaliado devidamente, ou estamos a fazer um somatório da avaliação das medidas?

P1: Depende muito dos programas. O FSE e o FEDER diferem muito, por exemplo. O FSE muitas vezes o que está a fazer é a refinar medidas. (...). Se estivermos a falar dos cursos profissionais, o POCH financia praticamente todos, só não financia Lisboa porque não é elegível. Portanto, quando formos avaliar, faz sentido avaliar todos e Lisboa não? Porquê tirar Lisboa? Simplesmente porque não é financiado pelos Fundos, mas pelo Orçamento de Estado?

Presumo que é importante avaliar as medidas, mas também a lógica do programa operacional...

P1: Temos essa preocupação. Ir ver do lado esquerdo da TdM até ao fim. O Programa Operacional está muitas vezes nas colunas iniciais, na definição dos critérios, etc. Mas claramente que em alguns casos pode ser mais descurada esta parte inicial, porque estamos mais centrados no fim da TdM, nos resultados. Mas depende muito das avaliações.

P2: Em teoria, avaliávamos isso nas avaliações de programa, porque nós temos avaliações temáticas, mas também de programa, que estão agora a decorrer – supostamente para serem intercalares, mas estão a ser finais, quase. A ideia de cobrir o programa estava contemplada no desenho do PGA. A verdade é que, por uma questão de rentabilização de esforços, o que fizemos quanto a essas avaliações de programa é que tentámos compô-las através da meta avaliação, quando já havia objetivos avaliados. Ou seja, quando se vai fazer a avaliação do POCH, avalia-se umas áreas, mas conta-se com a avaliação que já foi feita de algumas medidas. É verdade que isso pode retalhar a visão do programa.

P1: Depois reparem, dentro de um programa há situações muito distintas. No caso do POCH, há medidas que são geridas dentro do POCH, são os técnicos do POCH que analisam e aprovam candidaturas; e há outras medidas que são analisadas por outros organismos, quer servem de organismos intermédios. Há mesmo situações muito distintas dentro de um programa.

Outra questão: considerando todo o processo da AIC – desde a existência de dados, à escolha das técnicas, aos resultados e à sua comunicação – onde é que consideram existir maior potencial, e onde é que seria mais desejável haver inovação?

P1: Nós descansamos mais na AIC, porque achamos que ela já está mais consolidada no meio académico. Estamos mais focados na ABT. O que podia ser melhorado? Novas áreas de avaliação no Contrafactual. Já fizemos muitos exercícios de Contrafactual e sabemos como funciona. Agora é aproveitar e explorar outras áreas. Com as mesmas intervenções incluir análises distintas, ir progredindo. Por exemplo, avaliou-se a questão do sucesso e completar os cursos profissionais, porque não agora a empregabilidade? Introduzir variáveis de controlo mais robustas, por exemplo. Ir evoluindo. Portanto, mais intervenções, quer com mais variáveis, quer com mais variáveis de controlo.

P2: a CE organiza formações e foi apresentada uma metodologia de avaliação relativamente a territórios, em que não se supunha que houvesse grupos não intervencionados, e tentavam ver a amplitude do impacto. Alguns países iniciaram avaliações assim, com um exercício de Contrafactual. Para dizer que há possibilidades de contornar algumas limitações que o Contrafactual põe e que estão a ser desenvolvidas. Também, por outro lado, a possibilidade de articular com a TdM. Uma experiência que foi feita por um avaliador neste ciclo de programação foi a da construção da TdM para apoiar as variáveis e as desagregações pertinentes relativamente aos exercícios de Contrafactual. Inclusivamente, Contrafactual não só para aferir os impactos, mas para testar adesões, ou seja, em fases mais recuadas da cadeia, onde se articulam.

A AIC e ABT têm sido, efetivamente, usadas de forma complementar, mas se esse complementar tem sido independente; ou se tem havido uma complementaridade que adicione algo mais a ambas? A ligação entre as duas, que potencial traz?

P1: acho que esse trabalho pode ser melhorado, especialmente agora, com este trabalho de antecipação das TdM. Ao anteciparmos as TdM podemos melhor identificar essas complementaridades. Se for no mesmo estudo, tenho mais dúvidas que certezas, porque a ideia que temos, da nossa experiência, é que um Contrafactual apesar de responder a questões muito específicas, ao dar um número é muito apelativo. Depois, toda a ABT parece que se perde um pouco no contexto da avaliação. Parece que há uma aposta maior no Contrafactual, em detrimento de ABT. Ainda não temos bons exemplos... Temos bons exemplos só de Contrafactual, ainda não temos bons exemplos, no mesmo estudo, de combinados. Tenho muitas dúvidas de qual será a melhor solução. (...). Não estou certa se faz sentido no mesmo estudo ou em estudos distintos. Penso que uma pode comprometer outra. Embora, em teoria, parece que há casos em que faz sentido estarem juntos. Ainda não consigo fazer esse balanço, das vantagens de juntar ou separar. (...).

Referiu que gostaria de ter maior robustez nas variáveis de controlo, ou nas variáveis de resultado. Acha que a teoria conseguiria contribuir para esta robustez, pelo menos na seleção das variáveis de controlo, na sua definição?

P1: Acho que sim, que contribui para identificar variáveis relevantes e para conseguir de forma antecipada conseguir obtê-las. Mas isto não tem sido feito. O que tem sido feito é: o que é que existe? E montamos a avaliação com o que existe. Mas como disse, isto foi há 7 anos. Recuar 7 anos nestas áreas é bastante. A TdM pode ajudar em perceber porquê estas variáveis e não outras. Vai ter de haver sempre este equilíbrio entre o que há disponível e o que queremos fazer.

Há pouco falaram do IEJ. Eles têm o impacto médio global e o resultado por subgrupos de interesse ou heterogeneidades. Estas heterogeneidades podem ser explicadas ou não pela teoria...

P1: Exato, é mesmo isso. Muitas vezes, até podemos ter uma base de dados com informação, mas chegamos à conclusão que existe uma variável que é relevante e a mesma pode estar na base de dados de outro organismo, e que não possamos fazer esse cruzamento. Mas para isso temos de chegar à conclusão que é relevante. Temos de fazer este caminho.

P2: Gostava de referir uma coisa em que sentimos dificuldade e que são antigas. É o caso de construção de questionários, tradução da linguagem dos avaliadores em linguagem que as pessoas percebam, métodos de reporte, métodos de tratamento que testem as alegações causais que se fazem. Há uma série de requisitos que o trabalho académico exige, como apresentar R, como cuidado com afirmações que se fazem, mesmo com uso de dados quantitativos, e que sentimos que tínhamos muitos benefícios em que fossem transportados para o trabalho da avaliação. Como percebi que tinham essa ambição no vosso Guião, acho que devem cobrir isto.

