

AVALIAÇÃO DE POLÍTICAS PÚBLICAS E PROGRAMAS GOVERNAMENTAIS

Fundamentos e Modelos

**Evert Vedung
Luiz Pedone**

SOBRE O LIVRO

Avaliação, a ideia de sistematicamente aferir os efeitos das intervenções públicas e dos processos de implementação, tornou-se um ramo positivamente viável da análise contemporânea de políticas públicas. Adotada em todas as áreas do setor público, com a expectativa de que a avaliação possa suprir os tomadores de decisão (e seus oponentes) com argumentos baseados na melhor evidência empírica.

Adotando uma visão transdisciplinar de avaliação, o livro de Evert Vedung *"Public Policy and Program Evaluation"* é um tratado clássico no campo. O livro em português com Luiz Pedone baseia-se em um manuscrito bastante revisado, ampliado e modificado da primeira publicação em inglês.

O livro destina-se a matérias de avaliação em ciência política, educação, sociologia, psicologia, administração, serviço social, saúde pública, engenharia, ciências ambientais, estudos de defesa e disciplinas similares nas universidades. É relevante para cursos específicos dirigidos a servidores públicos.

AVALIAÇÃO DE POLÍTICAS PÚBLICAS
E
PROGRAMAS GOVERNAMENTAIS
FUNDAMENTOS E MODELOS

EVERT VEDUNG
LUIZ PEDONE

**AVALIAÇÃO DE POLÍTICAS PÚBLICAS
E
PROGRAMAS GOVERNAMENTAIS
FUNDAMENTOS E MODELOS**

EDITORA LUZES
COMUNICAÇÃO, ARTE & CULTURA
RIO DE JANEIRO
2021

Copyright © 2021 by
Evert Vedung
Luiz Pedone

EDITORES

Ubirajara Carvalho da Cruz
Shirley Santos da Cruz

CONSULTORIA

Prof^a Catarina Laboré Madeira Barreto Ferreira
Prof. Paulo Pereira Santos
Prof. Paulo Roberto Batista

REVISÃO

Autores

CAPA

Equipe Luzes

*É proibida a reprodução total ou parcial, por qualquer meio ou processo.
A violação de direitos autorais constitui crime (Código Penal, art. 184 e
Lei nº 6.895, de 7/12/1980), sujeitando-se a busca, apreensão indenizações
diversas (Lei nº 9.610/98).*

Os conteúdos desta obra são de responsabilidade de seus autores
Ficha Catalográfica na Editora

V414 Vedung, Evert e Pedone, Luiz
Avaliação de Políticas Públicas e Programas
Governamentais: Fundamentos e Modelos/Evert Vedung (1938-)
& Luiz Pedone (1948-)
Rio de Janeiro: Luzes - Comunicação, Arte & Cultura, 2021
326 p.

ISBN: 978-65992207-6-0

1 .Avaliação de Políticas Públicas 2. Ciência Política 3. Políticas
Públicas 4. Teoria da Intervenção 5. Instrumentos de Políticas
Públicas 6. Modelos de Avaliação. I Título

CDD 320

2021
Impresso no Brasil
Printed in Brasil

CONSELHO EDITORIAL

Prof. Marcio Rocha (INEST/UFF) - Conselheiro-Executivo

Prof. Adílio Jorge Marques (UFF)

Embaixador Alexandre Addor Neto (MRE)

Prof^a Beatriz Bissio (UFRJ)

Prof. Jorge Calvário dos Santos (ESG)

Prof. José Miguel Arias Neto (UEL)

Prof. Marcos Costa Lima (UFPE)

Prof. Sérgio Aguillar (UNESP/Marília)

Prof^a Patrícia de Oliveira Matos (PPGCA/UNIFA)

Prof. William de Sousa Moreira (EGN)

SUMÁRIO

Prefácio	Paulo Calmon	15
Prólogo	Evert Vedung	23
Preâmbulo	Luiz Pedone	29
Capítulo 1 – AVALIAÇÃO: UM IMÃ SEMÂNTICO		35
1.1	Avaliação definida	36
1.2	Avaliação não é aferição <i>ex ante</i>	37
1.3	Avaliação é uma aferição <i>cuidadosa</i>	38
1.4	Avaliação concerne às <i>intervenções governamentais</i>	40
1.5	Avaliação é focada no <i>conteúdo da decisão, administração, outputs e outcomes</i>	41
1.6	Avaliação é aferição de atividades em curso e acabadas	45
1.7	Avaliação é mais do que aferição de impactos planeja	46
1.8	Metas intervencionistas não são os únicos critérios de valor permitidos	46
1.9	Crerios de valor e mérito podem não ser necessariamente apenas <i>ex ante</i>	47
1.10	Avaliação deve ser útil	48
1.11	Modelos sistêmicos são dispositivos heurísticos	49
Capítulo 2 – INTERVENÇÃO, AVALIAÇÃO, FEEDBACK, META-AVALIAÇÃO		51
2.1	O Modelo multicamada de governança	51
2.2	Meta-avaliação e auditoria	61
Capítulo 3 – TEORIAS DA INTERVENÇÃO COMO INSTRUMENTO DE AVALIAÇÃO		59
3.1	O que é uma teoria de intervenção	59
3.2	Inserido na intervenção institucionalizada	60
3.3	O modelo estendido de sistemas como um esquema de análise da avaliação	61
3.4	O caso da Declaração de Combustíveis	64
3.5	Requisitos para boas reconstruções de teorias de intervenção brutas	69
3.6	Como a teoria de intervenção (refinada) pode ser usada	74

3.7 Teoria substantiva, teoria de implementação	77
3.8 Múltiplas teorias de intervenção	79
3.9 Limitações do método	80
Capítulo 4 – MODELOS DE AVALIAÇÃO: UM RESUMO	83
4.1 Modelos substantivos e modelos econômicos	83
4.2 Modelo de avaliação da União Europeia	87
Capítulo 5 – MODELOS DE AVALIAÇÃO SUBSTANTIVA	89
5.1 Modelo de alcance de metas	89
5.2 Prós e contras do modelo de alcance de metas	91
5.3 Modelo de efeitos colaterais	98
5.4 Avaliação livre de metas	110
5.5 Modelos de relevância	113
5.6 Avaliação orientado ao cliente (Destinatário)	115
5.7 Prós e contras do modelo de avaliação orientado ao cliente	119
5.8 Avaliação do <i>stakeholder</i>	119
5.9 Prós e contras da avaliação do <i>stakeholder</i>	124
5.10 Modelos de avaliação por colegiado: crítica dos pares, autoavaliação e combinações	125
Capítulo 6 – MODELOS ECONÔMICOS	129
6.1 Avaliação econômica inclui custos de intervenção	129
6.2 Avaliação econômica: produtividade	130
6.3 Avaliação econômica: eficiência	133
6.4 Pontos de referência	135
6.5 Sobre valor e critério	137
Capítulo 7 – ABORDAGEM DOS OITO PROBLEMAS EM AVALIAÇÃO DE POLÍTICAS PÚBLICAS	151
Capítulo 8 – PROPÓSITO BÁSICO, PROPÓSITOS SUPERIORES	157
8.1 Conceitos: o problema do propósito básico e dos propósitos superiores	157
8.2 <i>Accountability</i> : o agente presta conta, o principal afere	159
8.3 Quatro perspectivas de <i>Accountability</i>	161

8.4	Armadilhas sociais na avaliação de <i>Accountability</i>	163
8.5	Melhoria, Desenvolvimento	167
8.6	Avaliação para conhecimento básico	168
8.7	Propósitos estratégicos (“Vilas de Potemkim”)	168
Capítulo 9 – AUTOAVALIAÇÃO, AVALIAÇÃO INTERNA OU EXTERNA		171
9.1	Avaliação interna ou externa?	172
9.2	Prestação de contas	174
9.3	Melhoria formativa	175
9.4	Melhoria aditiva	177
9.5	Conhecimento Básico	177
Capítulo 10 – OBJETOS DA AVALIAÇÃO		179
10.1	Políticas, programas, componentes, projetos, elementos	180
10.2	Instrumentos de Políticas Públicas e Pacotes de Instrumentos de Políticas	180
10.3	Questionários Padronizados para Coleta de Dados	187
10.4	Gestão orientada a resultados ou a processos	189
10.5	Orientado a setores, orientado a áreas ou integrado (<i>mainstreamed</i>)	190
10.6	Governos Conjuntos/Co-Governos, parcerias (<i>governança colaborativa</i>) e subcontratação	192
Capítulo 11 – MONITORAMENTO		193
11.1	A abordagem de 9 passos de monitoramento	194
Capítulo 12 – MÉTODOS DE COLETA DE DADOS: QUATRO E APENAS QUATRO		201
12.1	Métodos Documentais	201
12.2	Métodos Audiovisuais	202
12.3	Métodos Interrogativos	204
12.4	Métodos de Observação	205
Capítulo 13 – EFEITOS: EXPERIMENTAÇÃO EM DOIS GRUPOS		209
13.1	Delineamentos para aferição de impacto	212
13.2	Experimentos com controles randomizados (ECR)	215

13.3 Efeito Hawthorne	219
13.4 Experimentos com Controles Pareados	222
13.5 Controles Combinados: O Delineamento de Séries Temporais de Controle	224
13.6 Ameaças à Validade Interna	225
13.7 Programas Públicos Deveriam Ser Experimentação Social?	227
Capítulo 14 – CONTROLES GENÉRICOS, REFLEXIVOS E CONTROLES-SOMBRA	237
14.1 Controles Genéricos	237
14.2 Controles Reflexivos	240
14.3 Delineamento de Séries-Temporais Interrompidas e Delineamento Antes-e-Depois para um Grupo	243
14.4 Controles-Sombra	247
Capítulo 15 – AVALIAÇÃO POR PROCESS TRACING E TEORIA DA IMPLEMENTAÇÃO	253
15.1 Avaliação por Process Tracing é Conduzida como Estudo de Caso	253
15.2 Abrindo a Caixa-Preta	254
15.3 Mecanismos	254
15.4 Process Tracing como Teoria da Intervenção	258
15.5 Sete Fatores Amplos que Podem Influenciar os Outcomes da Intervenção	259
15.6 Implementação 1: Fatores Efetivos na Intervenção	264
15.7 Implementação 2: Respostas dos Níveis Superior e Intermediário	270
15.8 Implementação 3: Respostas do Operador de Nível de Rua	274
15.9 Implementação 4: Respostas do Grupo-Alvo	276
15.10 Três Contextos que cercam o núcleo da implementação	279
15.11 Contexto 1: Formação da Intervenção e Sombra do Futuro	280
15.12 Contexto 2: Intervenções Paralelas, Outros Stakeholders, Cidadãos, Mercados	283
15.13 Contexto 3: Monitoramento e Avaliação Ex-Post	286
Referências	290

FIGURAS, TABELAS E CAIXAS DE TEXTO

FIGURAS

Figura 1.1 O Modelo Sistêmico Simples	41
Figura 1.2 O Modelo Sistêmico Adaptado para Avaliação de Intervenção Governamental	42
Figure 1.3 O Programa de Ajuda para Refugiados Afegãos	43
Figura 2,1 Avaliação no Modelo Multicamada de Governança	52
Figura 2.2 Meta-revisão como Meta-avaliação 1 e 2 no Modelo Multicamada de Governança	57
Figura 3.1 Modelo de Sistema com Administração, Output, Outcomes e seus Efeitos	62
Figura 3.2 Modelo de Sistema com Destinatários Intermediários e Finais	63
Figura 3.3 Modelo de Sistemas com Gestores e Operadores.	64
Figura 3.4 Um Primeiro Corte: a Teoria de Intervenção da Declaração de Combustível Reconstruída	67
Figura 4.1 Modelos Substantivos e Econômicos Exibidos em uma Lista simples como esquema tipo árvore	84
Figura 4.2 Modelos de Avaliação como galhos, ramos e brotos de uma árvore	85
Figura 4.3 Modelo de Avaliação da União Europeia	87
Figura 5.1 Avaliação de Alcance de Metas (avaliação de efetividade)	90
Figura 5.2 Modelo de Efeitos Colaterais	98
Figura 5.3 Efeitos Principais, Efeitos Nulos, Efeitos Perversos e Efeitos Colaterais	100

Figura 5.4 Avaliação dos Efeitos Colaterais com Efeitos Colaterais Especificados	108
Figura 5.5 Avaliação Livre de Metas	111
Figura 5.6 O Modelo de Relevância	113
Figura 5.7 Potenciais Stakeholders em Intervenções Públicas Locais	120
Figura 6.1 Produtividade	130
Figura 6.2 Eficiência como (1) Custo-Benefício e (2) Custo-Efetividade	135
Figura 6.3 Oito Problemas em Avaliação de Políticas Públicas (OPA) Problema # 7	138
Figura 6.4 Critério de Valoração nos Nove Modelos de Avaliação	141
Figura 6.5 Critérios de Valoração Adicionais para a Aferição de Intervenções Públicas	141
Figura 6.6 Critério de Avaliação para um Processo Completamente Democrático	147
Figura 7.1 Abordagem dos Oito Problemas em Avaliação de Políticas Públicas (EPA)	151
Figura 8.1 Duplo Papel na Cadeia de Representação Pública	164
Figura 10.1 Três Ferramentas de Políticas: Porretes, Cenouras e Sermões	180
Figura 10.2 Instrumentos de políticas afirmativos-negativos	185
Figura 13.1 O Problema do Impacto em Termos Contrafactuais	210

Figura 13.2 Delineamento de Pesquisa de Avaliação para Esclarecer Impactos	213
Figura 13.3 Experimentação Randomizada com medidas antes e depois do programa	215
Figura 13.4 Teoria da Situação do Campo de Conservação de Energia por Domicílio	223
Figura 13.5 Dois experimentos concordantes	224
Figura 14.1 Dois Projetos de Controles Reflexivos — Séries-Temporais Interrompidas e Antes-e-Depois para um Grupo — Como Quase Experimentos Amputados	242
Figura 14.2 Controles-Sombra	248
Figura 14.3 Controles-Sombra: Fontes de Dados e Questões Apropriadas	249
Figura 15.1 Process Tracing – Fatores de consolidação	262
Figura 15.2 Process Tracing – Fatores de consolidação graficamente	264

TABELAS

Tabela 5.1 Efeitos do programa sueco de subsídios para a conservação de energia em edifícios e processos industriais. 1974-1980	105
Tabela 10.1 Tipos de Objetos de Avaliação	179
Tabela 10.2 Abordagem de quatro passos para a gestão orientada aos resultados	189

CAIXAS DE TEXTO

Caixa 3.1 Os Termos “teoria de intervenção”, “teoria de programa” e relacionados	60
Caixa 3.2 Mecanismos generativos, premissas para a tomada de decisão	71
Caixa 3.3 Doutrina de interpretação de Popper: construindo uma posição que vale a pena atacar	74
Caixa 13.1 Profecias autorrealizáveis	220
Caixa 13.2 Delineamento da Série Temporal de Controle: Fatores a considerar e ameaças a evitar	226

PREFÁCIO

Paulo Calmon

Instituto de Ciência Política

Universidade de Brasília

“Avaliação de Políticas Públicas e Programas Governamentais: Fundamentos e Modelos” é um livro essencial, pois analisa temas importantes do campo da avaliação de políticas públicas de forma sistemática e atualizada.

Seus autores são renomados professores e pesquisadores com longa atuação na área de políticas públicas e avaliação de políticas programas. Professor Evert Vedung é uma referência internacional e autor de diversos livros e artigos acadêmicos de grande impacto. Sua vasta obra tem trazido diversas contribuições muito importantes para o campo da avaliação. Professor Emérito de Ciência Política da Uppsala University, um dos fundadores da Sociedade Sueca de Avaliação e mantenedor da Sociedade Europeia de Avaliação, ministrou aulas e foi pesquisador em algumas das mais prestigiadas universidades do mundo (Harvard, UT Austin, Linnaeus, Viena, Kyung Hee, Helsinki, Aalborg, entre outras). No Brasil, Professor Evert tem ministrado cursos na Universidade Federal Fluminense, na Universidade de Brasília, no Tribunal de Contas da União (TCU), na ENAP (Escola Nacional de Administração Pública), além de palestras em diversos órgãos federais, inclusive no Senado Federal, IPEA, Controladoria Geral da União, dentre outros. Professor Luiz Pedone é um respeitado professor de Ciência Política da Universidade Federal Fluminense, com atuação destacada na área de políticas públicas. Foi chefe do Departamento de Ciência Política da Universidade de Brasília, onde atuou como professor por quase três décadas. Engenheiro por formação, Mestre em Administração e PhD em Ciência Política pela Universidade de Massachusetts em Amherst, é autor de diversos artigos na área de políticas públicas, avaliação de programas e política comparada.

Além de todas as virtudes já citadas do livro e dos seus autores, é importante enfatizar que trata-se de um texto raro e que está sendo lançado em um momento muito oportuno, pois vem suprir uma lacuna importante na coleção de textos acadêmicos disponíveis em

língua portuguesa nas áreas de Ciência Política, Administração Pública e Gestão de Políticas Públicas. Atualmente, são raros os textos sobre avaliação de políticas públicas e programas governamentais disponíveis em língua portuguesa.

Nesse sentido, cabe indagar: Por que ao longo das últimas décadas a área da avaliação despertou relativamente pouco interesse no Brasil? Por que o campo da avaliação permaneceu tanto tempo numa posição secundária, a despeito de ser tão relevante para a gestão das políticas públicas?

A trajetória da construção do campo da avaliação no Brasil é idiossincrática. O interesse pelo tema surgiu de forma um tanto tardia. Apenas ao longo dos últimos anos é que tem surgido uma janela de oportunidade para inclusão da área de avaliação na agenda governamental. Em contraste, nos países pioneiros o tema foi incorporado no início dos anos 60 e ganhou maior aceitação no cotidiano da gestão pública na maioria das democracias contemporâneas. E muitos desses países já implementaram requisito legal exigindo que os principais programas governamentais sejam sujeitos a avaliações sistemáticas. Em função disso, formou-se um grande acervo de textos acadêmicos e relatórios técnicos sobre o tema. Mais do que isso, o campo da avaliação ganhou autonomia e identidade própria, o que é atestado pelo surgimento de diversas associações científicas e profissionais especializadas nessa área.

Se o campo da avaliação existe pelo menos desde as décadas de 1960 e 1970, por que que o Brasil está ingressando nessa área com quase meio século de atraso? A resposta não é simples. Certamente não se pode atribuir à ausência de recursos humanos capacitados no setor público brasileiro. Há várias escolas de administração pública instaladas no país desde os anos 60, além de centros de capacitação profissional à disposição dos servidores públicos. Também não se pode atribuir à falta de interesse em temas relacionados à área de gestão pública, já que a busca de mais eficiência e efetividade nas políticas públicas tem sido incorporada ao discurso de quase todos os governos ao longo dos últimos anos e suscitou diversas tentativas de reforma na administração pública.

Aqui vale uma observação. É curioso notar que, paradoxalmente a um certo desinteresse na área de avaliação, a área de planejamento

governamental recebeu grande atenção dos gestores de políticas públicas nas últimas décadas. Ao contrário do que ocorreu com a avaliação, o planejamento foi devidamente institucionalizado na administração pública brasileira, inclusive na própria Constituição Brasileira, muito em consequência da influência de economistas com orientação desenvolvimentista. O resultado disso é que, ao longo das últimas décadas, o país acumulou um vasto estoque de planos nacionais, regionais e setoriais que, pelo menos em tese, deveriam alinhar a atuação dos diversos órgãos governamentais e orientar as políticas públicas.

Cabe reconhecer que a avaliação de políticas e programas governamentais brasileiros não foi totalmente ignorada ao longo das últimas décadas. Organizações internacionais atuando no país avaliaram praticamente todos os programas apoiados por elas. Como exemplo, podemos citar diversos programas na área de saúde pública ligados ao SUS (Sistema Único de Saúde) e que foram apoiados por empréstimos do Banco Mundial (BIRD) e do Banco Interamericano de Desenvolvimento (BID). Esses programas (Reforsus, Vigisus, Proesf, dentre outros) foram sistematicamente avaliados nos seus mais diversos aspectos gerando diversos relatórios e publicações acadêmicas.

Além das organizações internacionais, alguns setores específicos dos governos federal e estadual desenvolveram sistemas avaliativos. Exemplo disso são os sistemas surgidos na área de educação pública e que foram concebidos e implementados por especialistas brasileiros da área educacional. Podemos citar os exemplos do Sistema de Avaliação da Pós-Graduação criado pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) em 1976 e o SAEB (Sistema de Avaliação da Educação Básica), estabelecido em meados dos anos 1990 pelo Ministério da Educação.

Essas iniciativas foram importantes e merecem ser reconhecidas. No entanto, elas ocorreram de forma isolada e não podem ser interpretadas como sinal de que práticas avaliativas estavam sendo devidamente incorporadas à agenda da administração pública brasileira. Para tentar compreender melhor o porquê do atraso na incorporação dessas práticas é interessante compreender como se deu o processo de institucionalização da avaliação em outros países. O caso dos Estados Unidos talvez tenha sido o mais emblemático e melhor documentado.

A institucionalização da avaliação nos Estados Unidos ocorre fundamentalmente no início da década de 1960 em decorrência da implantação de um amplo conjunto de novos programas sociais. A rápida expansão desses programas foi alvo de muitas críticas de diferentes grupos políticos que duvidavam da sua efetividade. Boa parte dessas dúvidas era alimentada por notícias sobre a ocorrência de abusos, fraudes e desperdício de recursos públicos. Diante desse contexto, o Congresso dos EUA passou a demandar que os novos programas fossem avaliados sistematicamente e, mais do que isso, passou a incluir na dotação orçamentária desses programas os recursos necessários para realização dessas avaliações. Em outras palavras, foi estabelecida e institucionalizada a demanda pela avaliação dos programas governamentais.

O súbito surgimento de uma ampla demanda por avaliações pegou todos de surpresa. O desafio era grande, já que havia poucos especialistas na área e a administração pública federal não estava preparada para atender essa demanda. Mas a pressão do Congresso e da sociedade, assim como a existência de recursos para realizar esses estudos, criou incentivos para que fosse desenvolvida a capacidade necessária para suprir essa demanda. Aos poucos essa capacidade foi sendo estabelecida, com o surgimento de novas estruturas organizacionais, a formação de uma comunidade de especialistas na área e o desenvolvimento de novos sistemas, novas técnicas e novas metodologias para avaliação de políticas públicas e programas governamentais. Em outras palavras, a função avaliação foi sendo institucionalizada na administração pública nos EUA.

A institucionalização da função avaliação nos EUA não foi um processo linear. Ocorreram alguns retrocessos, especialmente em períodos de crise econômica ou na presença de governos conservadores que estavam mais interessados na extinção dos programas governamentais do que no seu aperfeiçoamento. Paralelamente, como o governo naquele país possui uma estrutura complexa e altamente descentralizada, a avaliação foi sendo desenvolvida e implementada com base em diversos quadros de referência e para atender diferentes demandas específicas. Além das demandas estabelecidas no âmbito federal, havia outras demandas por parte dos estados e municípios, além dos chamados “think tanks”

- organizações não governamentais dedicadas às atividades de lobby e defesa de interesses em áreas específicas de políticas públicas.

Processos mais ou menos similares ocorreram em outros países desenvolvidos, embora cada um tenha suas próprias especificidades. O denominador comum entre eles é o fato de que a avaliação se institucionalizou a partir de uma combinação de dois fatores: (i) a expansão dos programas governamentais objetivando o estabelecimento de um Estado de Bem Estar Social; (ii) as demandas da sociedade por maior transparência e maior efetividade nas políticas públicas e programas governamentais.

Até muito recentemente esses dois fatores não estavam presentes no Brasil. De 1960 até o início dos anos 80 o Brasil estava sob tutela de um regime autoritário e o Congresso Nacional tinha uma influência muito limitada nas decisões sobre políticas públicas. Ao longo desse período, as decisões eram tomadas quase que exclusivamente no âmbito do Poder Executivo, e a ênfase era na formulação e gestão de planos nacionais e setoriais. A demanda por maior transparência e efetividade era restrita e insuficiente para impulsionar a estruturação da função avaliação no governo federal.

No período que vai de meados dos anos 80 até o início dos anos 90, o Brasil se redemocratiza, o Congresso Nacional ganha maior relevância nas decisões sobre políticas públicas e uma nova geração de programas governamentais começava a ser concebida. Mas essas mudanças tiveram pouco efeito concreto na demanda por avaliação. O país mergulhou em uma crise econômica profunda, provocada pelo desequilíbrio estrutural do balanço de pagamentos, e que gerou altas taxas de inflação, crise fiscal e a desorganização da gestão pública. Ao longo desse período, também chamado de “A Década Perdida”, a busca da estabilidade macroeconômica passava a ser prioridade absoluta e, mais uma vez, as práticas avaliativas foram relegadas a um segundo plano.

Apenas a partir do início desse novo século esse cenário começa a se alterar e a demanda por avaliação se consolida e expande. Vários fatores têm contribuído para isso. Primeiramente, a exemplo do que ocorreu em outros países nos anos 60, houve uma expansão considerável dos programas governamentais, na medida em que o país se engajou em um projeto de desenvolvimento inclusivo. O caso

mais emblemático talvez seja o Programa Bolsa Família, estabelecido em 2003 e que hoje atende quase 20% da população brasileira. Paralelamente, a demanda por efetividade e transparência também se eleva, impulsionada por diversos escândalos envolvendo recursos públicos.

Um bom exemplo de resposta institucional à crescente demanda por avaliação é o que ocorreu no Ministério do Desenvolvimento Social (MDS), responsável por administrar boa parte da nova geração de programas sociais, inclusive o Programa Bolsa Família. Lá foi estabelecida uma estrutura avaliativa especializada, a Secretaria de Avaliação e Gestão da Informação (SAGI), que fomentou o estabelecimento de uma ampla rede interdisciplinar de especialistas em avaliação. Essa rede contou com apoio de vários parceiros, inclusive o IBGE (Instituto Brasileiro de Geografia e Estatística), o CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e diversas instituições de ensino e pesquisa federais e estaduais. Como resultado, entre 2013 e 2016 foi desenvolvido um amplo conjunto de avaliações sobre diversos aspectos dessa nova geração de programas sociais, totalizando quase 200 trabalhos técnicos nas quatro grandes áreas programáticas de atuação do MDS: Transferência de Renda, Assistência Social, Segurança Alimentar e Inclusão Produtiva.

Processo mais ou menos similar ao ocorrido no MDS, aconteceu em algumas outras unidades do governo federal que começam a incorporar a avaliação as suas atividades. Esse movimento não é coordenado, mas tem contribuído gradativamente para a institucionalização da avaliação no setor público brasileiro. Há ainda muito a ser feito e, com certeza, esse processo de institucionalização não é linear e ainda pode sofrer revezes. No entanto, há fortes indícios de que os incentivos criados pelo novo contexto político e pela crescente demanda por mais transparência e efetividades nos programas governamentais continuarão impulsionando a disseminação e consolidação das práticas avaliativas no setor público Brasil.

Dentre os elementos essenciais para a sustentação desse impulso pela expansão e institucionalização da função avaliação na administração pública brasileira, destaca-se o desenvolvimento de uma ampla comunidade de pesquisadores especializados no tema. Para tanto, é fundamental a disponibilização de textos didáticos de

alto nível que auxiliem professores e alunos a compreender melhor os conceitos, as metodologias e as estratégias de pesquisa que foram desenvolvidas ao longo dos últimos 50 anos e que hoje formam o campo da avaliação de políticas públicas e programas governamentais.

A obra produzida pelos professores Vedung e Pedone tem exatamente essas características. É um texto que proporciona, de forma clara e didática, os elementos necessários para que o leitor conheça o campo da avaliação de políticas e programas e possa contribuir para produção de pesquisas e estudos que contribuam para dar maior efetividade e transparência às políticas públicas e aos programas governamentais. É, portanto, um texto que interessa muito aos estudantes, professores e pesquisadores das áreas de Ciência Política, Administração Pública, Gestão de Políticas Públicas, Economia, Ciências Sociais, Direito e Contabilidade. Mas creio também que será de grande valia para todo cidadão disposto a contribuir para a construção de uma sociedade mais democrática, mais justa e com políticas públicas efetivas, capazes de atender aos reais anseios e necessidades da população.

Brasília, novembro 2018

Paulo Calmon

PRÓLOGO PARA A EDIÇÃO REVISADA EM PORTUGUÊS

Evert Vedung

Uppsala University

Central para a tradução e publicação do meu manuscrito sobre **avaliação de políticas públicas** no Brasil foi o meu bom amigo, professor Luiz Pedone. Com entusiasmo, ele apoiou minha sugestão de ter uma versão fortemente revisada do meu livro “Public Policy and Program Evaluation”, de 1997, traduzida para o português. Ele concordou animadamente em criar um grupo de tradução entre nossos alunos de avaliação da Universidade Federal Fluminense (UFF) em Niterói, no Rio de Janeiro, e supervisionar seus trabalhos no Laboratório *Defesa & Política[s]*. Ao participar de nosso minicurso a partir de 2013, alguns dos alunos familiarizaram-se com a terminologia específica em inglês e adquiriram o conhecimento necessário e o interesse em avaliação para poder produzir uma interpretação razoavelmente boa para o português, o que pode melhorar depois em inúmeras revisões em português.

Para avançar este projeto, Pedone convidou-me para ir ao Brasil, pediu e recebeu recursos para cobrir os custos da minha viagem Estocolmo-Rio; praticou todos os trâmites e preparativos necessários no Instituto de Estudos Estratégicos (INEST) da UFF para os cursos anuais de 2013 a 2017, tendo recrutado participantes capazes e inteligentes. Os cursos foram bem. Curiosamente, alguns participantes do primeiro lote participaram do curso pela segunda vez também. Nossa exigência de que os alunos deveriam compor os trabalhos em inglês encontrou dificuldades lingüísticas, mas foi geralmente aplaudida pelos estudantes de língua portuguesa. Surpreendente para mim foi didática que Pedone imprimiu aos cursos: as palestras matinais (aula expositiva-teórica) eram seguidas, depois do almoço, por uma tutoria intensiva para todo o grupo até o final da tarde. Levei algum tempo para perceber que esse excelente procedimento de supervisão não era peculiar ao Brasil, mas peculiar ao professor Pedone.

Logo no início enviei-lhe a maior parte da versão já revista do meu livro de 1997 para tradução. Atualizar os dois últimos capítulos do livro original - ou talvez inteiramente excluí-los - era uma tarefa que me atormentava há muito tempo, o que exigiu que Pedone repetidamente me lembrasse e alertasse a necessidade. No final, conseguimos terminar a tradução e remunerar os alunos.

A tradução para o português foi feita principalmente por Mariana Guimarães Alves da Silveira e Lucas Peixoto Pinheiro da Silva. Em menor grau, Leandro Augusto Pinotti, Urias da Silva Fernandes Jr e Pedro Henrique Peres Suzano e Silva também participaram. Eles fizeram uma primeira versão crua que Pedone revisou e entregou para Lucas Peixoto e Mariana Guimarães para um segundo e terceiro refinamento. Mariana, Lucas e Pedone fizeram novas revisões do capítulo 1 até o final. E Pedone fez a revisão final de todos os quinze capítulos, ajustando as notas, referências e bibliografia às especificações da Associação Brasileira de Normas Técnicas (ABNT). Sem as habilidades de liderança, persistência e posição de Luiz Pedone no INEST e em algumas fundações de pesquisa brasileiras, somadas à devoção e ambição dos estudantes acima mencionados, a tradução nunca teria visto a luz do dia. – No INEST-UFF, vou também agradecer o apoio de Eurico de Lima Figueiredo, Marcio Rocha e Barbara Dutra Lameira Ottero.

Além disso, Luiz e sua mulher Maité Baena foram anfitriões maravilhosamente generosos e vivazes, amigos leais e de alto astral durante todas as minhas visitas ao Rio.

Pedone foi fundamental para minha vinda ao Rio e à UFF em primeiro lugar para ensinar avaliação de políticas públicas. Mas as circunstâncias favoráveis a esse desenvolvimento inesperado foram criadas por meu colega e querido amigo Mikael Román, de quem fui tutor oficial para sua tese de doutorado de 1998 sobre a implementação do Tratado de Cooperação Amazônica. Román mais tarde foi empregado pela Agência Sueca para Análise de Políticas de Crescimento (Tillväxtanalys) e colocado na Embaixada da Suécia em Brasília, Mikael me pediu para visitá-lo e dar uma olhada mais de perto na famosa capital do Brasil como um monumento notável da política urbana no auge da modernidade nos anos 50. Em estreita colaboração com o professor Paulo du Pin Calmon, renomado cientista político da

Universidade de Brasília (UnB), Mikael criou um programa incrível para mim, incluindo conversas com colegas de ciência política como Wellington Almeida, um seminário na UnB, uma visita guiada e conversa com Carmen Meireles, Selma Maria Hayakawa Cunha Serpa e Glória Maria Merola da Costa Bastos do Tribunal de Contas da União (TCU), uma curta sessão no TCU e, como grand finale, uma visita pessoal ao Rio de Janeiro com Ipanema, Pão de Açúcar e Cristo Redentor. Foi lá, em 4 de fevereiro de 2012, no restaurante La Fiorentina, no Leme, final de Copacabana, em um encontro com Pedone e sua mulher Maité, instigado por um telefonema improvisado de Mikael para Luiz e por sugestão igualmente improvisada do último. que nós dois decidimos iniciar nossa cooperação acadêmica.

Nos anos que se seguiram, Mikael me ajudou e me inspirou de várias maneiras em Brasília e no Brasil. Ele me levou para a Cidade Administrativa de Niemeyer em Belo Horizonte, Ouro Preto, Mariana e Congonhas com Alejadinho no Estado de Minas Gerais; para Pirenópolis, em Goiás, com suas cachoeiras e arquitetura histórica; a Gramado, a Cascata do Caracol, a vinícola Lidio Carraro e outras maravilhas do Rio Grande do Sul; e para o Amazonas e Manaus, a montante do imenso Rio Negro, para Santarém e para o fundo das florestas tropicais ao longo do Rio Tapajós.

Em 2015, em nome de sua agência, Mikael começou a pressionar pelo aprimoramento do setor público como uma área promissora de inovações para a co-criação sueco-brasileira. Em 17 de novembro, ele organizou uma reunião com a direção da Escola Nacional de Administração Pública (ENAP) para descobrir formas de cooperação. Entre os presentes estavam Luis Henrique D’Andrea, diretor de cooperação internacional da ENAP (que havia sido aluno de Luiz Pedone na UnB), e Maria Stela Reis, diretora de formação profissional. Depois, Mikael, o embaixador sueco Per-Arne Hjelmhorn e eu executamos um painel público sobre as possibilidades e os limites da avaliação como um instrumento para melhorias e inovações do setor público, presidido por Paulo Sérgio de Carvalho, presidente da escola.

No ano seguinte, Mikael havia aperfeiçoado sua ideia na chamada “Semana de Inovação Suécia-Brasil” em Brasília e em várias outras cidades importantes. A questão delicada da avaliação da política de inovação seria apresentada como um novo tema deste evento,

centrado em torno do seu diretor de Estocolmo, Enrico Deiacó, e envolvendo também eu. Novamente na ENAP, uma Mesa Redonda foi organizada em novembro. Enrico falou vigorosamente sobre a abordagem econômica e quantitativa da avaliação da política de inovação, com dados contrafactuais, ou seja, a abordagem da regularidade à causalidade. Por outro lado, concentrei-me no rastreamento detalhado de um processo específico de implementação de políticas de inovação, sem um caso contrafactual, mas onde os avaliadores acompanham a implementação passo a passo durante um longo período de tempo para mapear a série de mecanismos que conectam partes da intervenção com seus processos, resultados não intencionais e surpreendentes, isto é, a abordagem do mecanismo à causalidade.

Como resultado das semanas de inovação de Mikael, dois valiosos contatos se seguiram. O professor Bernardo Kipnis, do Centro de Estudos Avançados de Governo e Administração (CEAG), confirmou que me queria como parceiro no processo de rastreamento em sua avaliação contínua planejada da grande iniciativa “Adolescentes em conflito com lei” dentro do grande projeto SINASE. No final daquele mês, um segundo contato foi estabelecido. Paulo Calmon me ajudou a executar uma “AULA MAGNA” na Escuela Interamericana de Administración e Políticas Públicas, EIAPP, realizada este ano na ENAP. Francisco Gaetani, então vice-ministro do Meio Ambiente e não totalmente desconhecido para mim, foi o orador principal. Depois do seu discurso, e depois de emitir algumas palavras amigáveis para mim, ele se desculpou e saiu. Mas no final da tarde ele ligou várias vezes e pudemos marcar uma reunião incluindo também Mikael Román às 9h30 da noite em um restaurante. Para minha surpresa, ele conhecia minhas contribuições para a literatura de instrumentos de políticas públicas no livro “Carrots-Sticks and Sermons”. E ele queria cooperação. Tendo tentado por meses estabelecer uma reunião com Gaetani em questões ambientais comuns, Mikael Román ficou satisfeito. De minha parte, isso levou a minha oferta de um curso ENAP de uma semana em 2017 e um convite para voltar em novembro de 2018.

De sua posição de professor e personalidade influente da UnB no CEAG, Paulo Calmon e sua enorme rede de antigos e atuais alunos, colegas e funcionários públicos foram de uma ajuda maravilhosa para

mim em Brasília e no Brasil. Na verdade, ele atuou como meu principal empreendedor brasileiro. Eu já mencionei o evento do EIAPP e o tremendo programa em minha primeira visita em 2012. Isso continuou ao longo dos anos. Além de ter me acompanhado nas compras pessoais, ele, incansavelmente, direta ou indiretamente, organizou seminários, almoços, jantares, palestras de café, visitas, visitas turísticas e outros eventos para mim com pessoas no Senado Federal (Ivana Trombka, Fernando Boaroto Menequin); Câmara dos Deputados (Fábio de Barros Correia Gomes); Secretaria de Orçamento Federal (Fernando Sertã Meressi, César Augusto Assis Mascarenhas de Oliveira); Controladoria Geral da União CGU (Hugo Antônio de Azevedo Lousã, Alexandre Gomide Lemos); Instituto de Pesquisa Econômica Aplicada IPEA (Jessé de Souza, Sandra Silvia Paulsen, Alexandre Gomide); Controladoria Geral do Distrito Federal CGDF (Henrique Morães Ziller); Universidade Católica de Brasília (Marcelo Fiche); UnB e UnB-CEAG (Igor da Silva Moraes; Adriano Martins Juras, Paulo Asafe C Spinola e outros alunos de pós-graduação em ciência política); UnB em Planaltina (Luiz Guilherme de Oliveira); UnB Instituto de Psicologia (Sheila Giardini Murta, Nadia Prazeres Pinheiro Carozzo, Ingrid Gomez Abdala)

Graças a ele, ministrei um curso de avaliação de intervenção pública por ano de 2013 a 2016, para acadêmicos da unb, auditores, etc. do Tribunal de Contas da União e outros funcionários públicos federais. Os mais envolvidos do lado do TCU foram Dagomar Henriques Lima, Selma M H Cunha Serpa e Maridel Piloto de Noronha.

Basicamente, os cursos foram semanais, com aulas das 9 da manhã até as 12:30 ou 1:00 da tarde. Todo dia de curso, ao longo de todos os quatro anos, Paulo me levou no seu carro do meu hotel para o local do curso e de volta. Ele esteve presente nos cursos, apresentando-me, informando os participantes sobre os próximos eventos, fazendo perguntas pertinentes e fornecendo comentários incisivos. Ele tem sido um apoio acadêmico tremendamente importante para mim durante as minhas aventuras no Brasil.

Além do acima, menciono Marcia Paterno Joppert, participante do primeiro lote do meu curso de 2013, que em 2014 me convidou para realizar um curso de treinamento de um dia inteiro na Rede Brasileiro de Monitoramento e Avaliação em Porto Alegre e depois, em 2015, para um dia maravilhoso em sua segunda casa em Pirenópolis;

Katya M N Calmon, amiga e provavelmente a primeira brasileira a usar meu livro de avaliação, agora traduzido; Rafael Ferrari, participante do curso, amigo, em Estocolmo e Brasília; Pedro Lusz, amigo, aficionado do Clube de Choro de Brasília; Loreley Rios, pela ajuda com cópias na embaixada sueca; Mariana Prisila Maculan Sodré, TCU, ansiosa participante do curso, em sua casa sediando duas festas inesquecíveis de confraternização pré-natal para pessoas do TCU, inclusive eu.

A liderança e a equipe do Instituto de Habitação e Pesquisa Urbana da Universidade de Uppsala (IBF) têm me apoiado muito, apesar do meu status de aposentado. Kerstin Larsson acelerou com competência as faturas e as questões tributárias para estrangeiros em torno de uma doação da Fundação Stint, que durante 2014-2015 financiou parcialmente a estada do professor Luiz Pedone na Suécia e minha viagem ao Rio e à Universidade Federal Fluminense. A Prefekt Irene Molina, diretora d IBF, tem apoiado fortemente minhas visitas recorrentes à universidade mencionada em Niterói/Rio de Janeiro. Christina Kjerrman-Meyer e Carita Ytterberg ajudaram na minha home page da IBF; Ulrika Wahlberg com e-mail e migração para um novo sistema de home page; Camilla Scheinert com questões de informação geral, questões de bibliografia e DiVA; e Peter Knutar com a função de e-mail.

O engenheiro florestal Tage Vedung transformou as minhas figuras desenhadas à mão em desenhos em aplicativos para computadores no manuscrito em inglês. Agora ajudou-me a selecionar e adquirir diversas parafernálias de computador e instalou novos programas e novas versões de Windows. Um ditado da saga islandesa de Grettis prega a importância de Tage para mim: "Indefesas estão suas costas sem um irmão por trás dela."¹ Assim eu sinto pelo meu irmão Tage; sua ajuda é indispensável.

Recebemos doações das seguintes fundações: em 2013 e 2014 da Faperj (Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro) e da UFF; e em 2015 da Fundação Sueca para Cooperação Internacional em Pesquisa e Ensino Superior (Stint), Estocolmo. Neste último caso, meu bom amigo e colega Nils Hertting, do IBF, foi de grande ajuda.

1. Islandês: "Bjerr er hvern á bakinu nema sér bróður eigi." Sueco: "Barr är ens rygg utan en broder bakom den". Inglês: "Bare is one's back without a brother behind it."

Quero estender meus calorosos e profundos agradecimentos a todas as pessoas e instituições mencionadas acima. Em particular quero agradecer: à UFF, aos alunos / tradutores Mariana Guimarães Alves da Silveira, Lucas Peixoto Pinheiro da Silva, Leandro Augusto Pinotti, Urias da Silva Fernandes Jr e Pedro Henrique Peres Suzano e Silva; e Márcio Rocha como responsável pela publicação do presente livro; da UnB, Bernardo Kipnis e Nair Aguiar de Miranda; do TCU, Dagomar Henriques Lima, Glória M M C Bastos, Selma M H Cunha Serpa, Maridel Piloto de Noronha; da ENAP, Francisco Gaetani, Maria Stela Reis, Cláudia Antico, Iara Alves, Pollyanna Rodrigues Costa, Antônio Claret Campos Filho e Armando Armorim Simões, e por último, mas não menos importante, Paulo Calmon, Luiz Pedone e Mikael Román.

Fålhagen, Uppsala, Agosto 2018

Evert Vedung

PREÂMBULO

Luiz Pedone

Instituto de Estudos Estratégicos

Universidade Federal Fluminense

Evert Vedung escreveu um livro clássico em sueco “Utvärdering i politik och förvaltning” (*Avaliação em política e gestão*) – publicado inicialmente em 1991– como resultado de suas pesquisas inicialmente sobre instrumentos de políticas de energia. Mais tarde, sua pesquisa mudou gradualmente para avaliar *o como e o porquê* da ação governamental sueca na política energética nos anos de crise entre 1973 e 1981. Vedung desenvolveu e aperfeiçoou seus métodos de avaliação em outras pesquisas, mas foi sua passagem pela University of Texas at Austin que o colocou em contato com a American Evaluation Association. Ali tomou conhecimento da “fermentação” que ocorria naquela comunidade epistêmica através de sua participação nos encontros anuais da AEA. Ao final dos anos Oitenta, Vedung já tinha o manuscrito pronto para ser publicado em 1991 no idioma sueco. O livro ultrapassou rapidamente as fronteiras da Uppsala University sendo adotado em inúmeros programas de ciência política, de governo e administração nas universidades nórdicas e europeias. Foi também muito utilizado em consultorias, contratos de pesquisa em avaliações dos programas de várias agências governamentais suecas de planejamento, de construção civil, auditoria e assuntos culturais.

Os conhecimentos de Vedung não somente se propagaram nos países escandinavos contíguos à Suécia. Durante suas andanças como professor visitante, pesquisador visitante e leitor pela Espanha, Áustria, Dinamarca, Coreia do Sul, Polônia, Finlândia e Estados Unidos, Vedung preparou um livro que não foi uma “mera tradução do meu livro sueco de 1991, mas um produto inteiramente revisado”, como ele mesmo diz no seu prefácio à edição inglesa (1997, p. xiv). A edição deste livro pela Transaction Publishers (New Jersey, EUA e Londres, Reino Unido) tornou-se um cânone em avaliação de políticas públicas. Em 1996 o Instituto Nacional de Servicios Sociales da Espanha o traduziu para utilização nas suas atividades.

Nesta edição em português um processo similar ocorreu. Em 2012, Vedung e eu estabelecemos um programa de cooperação acadêmica que consistiu na sua vinda à Universidade Federal Fluminense/Instituto de Estudos Estratégicos como professor visitante para lecionar em conjunto cinco cursos intensivos de Avaliação de Políticas Públicas, entre 2013 e 2017. Neste mesmo período estive seis vezes na Suécia como Pesquisador Sênior, algumas como bolsista da CAPES. Este foi um resultado direto da criação sob minha coordenação, do Grupo de Pesquisa Defesa, C&T e Política Internacional registrado junto ao CNPq em 2012, o qual, com o nome prático de Laboratório *Defesa&Política[s]*, promoveu os Cursos de Avaliação de Políticas Públicas registrados na Pró-Reitoria de Extensão da UFF, aprovados pelo Departamento de Estudos Estratégicos e Relações Internacionais e pelo Programa de Pós-Graduação em Estudos Estratégicos (PPGEST) do Instituto de Estudos Estratégicos, com o apoio de Thiago Rodrigues, chefe de departamento (2012-2017), de Vagner Camilo Alves, coordenador do PPGEST (2009-2017), e de Eurico Figueiredo, diretor do INEST-UFF. Formaram-se quase sessenta alunos nos cursos oferecidos naqueles anos.

A ideia de ter um livro básico para o curso foi logo aventada e uma das opções era traduzir o seu livro clássico publicado em inglês em 1997. Mas, Evert Vedung novamente não só reescreveu como ampliou o manuscrito do livro quase que inteiramente. Capítulos inteiros foram remanejados, tópicos específicos acrescentados. O tema da teoria da intervenção foi antecipado para o início do livro e caracterizado especificamente como uma ferramenta de avaliação, um fundamento para a subdisciplina. Os modelos de avaliação foram expandidos e consubstanciados em três novos capítulos. A caracterização da intervenção pública foi revista para abordar não só os instrumentos de políticas públicas, mas toda uma nova discussão sobre gestão orientada a resultados ou orientada a processos, orientada a setores, ou áreas governamentais específicas, assim como tema da governança colaborativa e parcerias entre níveis de governo.

Os dois últimos capítulos do livro original (*14. Merit Criteria and Performance Standards*; e *15. Uses and Utility of Evaluation*) foram

diluídos em outros capítulos do manuscrito revisado. O antigo capítulo 13 (*Process Evaluation and Implementation Theory*) sofreu uma grande mudança. Evert Vedung introduziu no novo capítulo 15 todos os conceitos contemporâneos de *Process Tracing* desenvolvidos nos últimos quinze anos e os associou com a *Teoria da Implementação*, depois de muitas discussões conjuntas. Trata-se de uma abordagem inovadora em avaliação de políticas públicas por *process tracing* examinando os elementos e fatores causais que se interconectam no processo de implementação. Embora tenha sido enviado bem depois dos primeiros capítulos originais, valeu a espera, porque este capítulo serviu como o coroamento de todo o trabalho.

Avaliação de políticas públicas e programas governamentais é uma atividade interdisciplinar que pode levantar controvérsias. Uma dimensão do conflito é considerar a avaliação como um experimento aleatório controlado e conduzido por acadêmicos para tornar políticas públicas mais científicas **vs.** considerar a avaliação como um processo dialógico participatório realizado pelos cidadãos comuns e pelos interessados no lugar dos acadêmicos para tornar políticas públicas mais democráticas. Outra dimensão do conflito refere-se ao uso das avaliações: deveriam elas ser para controle, “accountability” (responsabilização) e, portanto, independente dos que estão sendo avaliados **vs.** deveriam ser utilizadas para aperfeiçoamento e promoção, portanto, mais conectado aos que estão sendo avaliados? De qualquer modo, o campo da avaliação abriga controvérsias e parece mudar com o tempo em torno destas e outras dimensões, ou como diz Evert Vedung em sua metáfora – “em ondas de avaliação”.

A tradução inicial ficou a cargo de alunos excepcionais ligados ao *Laboratório Defesa&Política[s]*: Mariana Guimarães Alves da Silveira, Lucas Peixoto Pinheiro da Silva, Leandro Augusto Pinotti, Urias da Silva Fernandes Junior e Pedro Henrique Peres Suzano e Silva. As primeiras versões ficaram prontas em fins de 2015. Mas o trabalho mais árduo foi revisar parágrafo a parágrafo de cada capítulo, comparando com o original, trabalho este realizado por mim com uma releitura atenta de Mariana Guimarães e Lucas Pinheiro. Particularmente o trabalho de adequar a linguagem para os termos comumente utilizados em ciência política, políticas públicas,

administração pública no Brasil foi muito intenso, mas gratificante.

Evert Vedung fez uso de toda uma nova bibliografia inexistente em 1997. Isto implicou em uma extensa garimpagem eletrônica para obter as referências completas dos livros, artigos, documentos, teses, relatórios de pesquisa publicados depois daquela data e, obviamente, adequá-las às normas da ABNT. No final do livro apresentamos um complemento de Periódicos e Associações Profissionais de Políticas Públicas no Brasil, mostrando que, em 35 anos desde a redemocratização, o campo evoluiu significativamente entre nós. Um dos aspectos mais gratificantes da vida acadêmica é ampliar o alcance do conhecimento e ver a disseminação dos fundamentos propagarem-se e, para atingir este objetivo, somente alguém generoso como Evert Vedung poderia ter-me oferecido a coautoria deste trabalho agora publicado em português.

Alguns colegas e instituições foram fundamentais para tornar este livro realidade. Agradecemos o apoio editorial do Inest/UFF através do seu diretor, professor Titular Emérito Eurico de Lima Figueiredo, e do diretor do Núcleo de Estudos Estratégicos Avançados (NEA), professor Márcio Rocha. A cooperação acadêmica foi viabilizada por bolsas e custeio para todos os envolvidos pela FAPERJ (Fundação Carlos Chagas de Amparo à Pesquisa do Estado do Rio de Janeiro), pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), pela STINT (The Swedish Foundation for International Cooperation for Research and Higher Education), e pela PROPPi (Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação da UFF) ao longo destes anos. Pedimos ao nosso amigo e colega, Paulo Calmon, atual diretor do Instituto de Ciência Política da Universidade de Brasília, para participar da arguição dos trabalhos finais do Curso em 2016. Paulo Calmon, um “scholar” brilhante, é responsável em grande parte pelo desenvolvimento e consolidação do ensino e pesquisa em políticas públicas no Brasil. Depois disso, o convite para escrever o prefácio à edição brasileira foi natural.

Este livro destina-se a suprir as necessidades de livro básico de fundamentos e modelos para disciplinas de graduação e de pós-graduação (*lato sensu*, profissional e *stricto sensu*) em avaliação e análise de políticas públicas nos cursos de Ciência Política, Políticas Públicas, Relações Internacionais, Administração Pública e outras

áreas do conhecimento em ciências humanas e em ciências sociais aplicadas (Educação, Ciências Sociais, Sociologia, Serviço Social, Economia e Direito), cuja proliferação cruzada com políticas públicas pode-se notar nos programas de pós-graduação recentemente avaliados pela CAPES nestas áreas. Destina-se também a suprir as necessidades dos especialistas e técnicos em avaliação em carreiras mais contemporâneas dirigidas aos estudos e avaliações de políticas e programas em Defesa e Segurança, Ciência-Pesquisa-Tecnologia-Inovação, Gestão do Setor Público, Gestão de Políticas Públicas, Saúde Pública, Biotecnologia, Transporte, Comunicações, Meio Ambiente, Energia, Gestão do Trabalho e Recursos Humanos, Desenvolvimento Urbano, Desenvolvimento Social, Drogas e Narcotráfico, Crime Organizado, Migrações e Terrorismo, dentre outros existentes hoje no Brasil e no mundo.

Espera-se que este livro contribua para ampliar o conhecimento e permitir análises e avaliações mais efetivas nas atividades de professores, pesquisadores, estudantes de avaliação de programas e projetos governamentais nas instituições de ensino superior e de pesquisa públicas e privadas, incluindo as escolas militares superiores. Por fim, deve contribuir, da mesma forma, nas análises e avaliações executadas por profissionais de controle e avaliação nos mais diversos setores da administração pública, e nos institutos de controle e avaliação em diversos níveis de governo e na sociedade em geral.

Niterói-Rio de Janeiro, novembro de 2018

Luiz Pedone

AVALIAÇÃO: UM ÍMÃ SEMÂNTICO

Isso tem me causado um grande problema e ainda me causa um grande problema: perceber que *o nome que as coisas têm* é indescritivelmente mais importante do que o que elas são.

**Friedrich Nietzsche, *The Gay Science*
(*Die fröhliche Wissenschaft*) (2001), p. 69**

Apesar de muitas tentativas, nenhuma definição satisfatória de jardim e jardinagem foi encontrada; todas as definições existentes deixam uma grande área de incerteza sobre o que é de onde. Nós simplesmente não sabemos o que exatamente jardim e jardinagem são. Usar estes conceitos seria, portanto, intelectualmente irresponsável, e, na verdade, no caso de jardim seria ainda mais. Tu não jardinarás. Q.E.D.

**Lezek Kolakowski, *The General Theory of Not-Gardening*,
em *Modernity on Endless Trial* (1990)**

Na Antiguidade clássica, estudiosos eram chamados à corte para se tornarem conselheiros do príncipe. Aristóteles foi contratado pelo rei Felipe da Macedônia como professor de estadismo do jovem Alexandre. Durante o cerco de Siracusa, os legionários romanos foram forçados a se protegerem de espelhos ustórios de Arquimedes e de catapultas. A tendência continuou nos Estados-nações da Europa. Atendendo ao pedido do rei Cristiano IV, o proeminente astrônomo Tycho Brahe se estabeleceu na corte de Copenhague para ler o horóscopo do monarca a fim de auxiliar o rei na composição de sua política externa. Membros da família Nobel, em seus esforços para inventar e suprir o Czar com armas modernas fizeram vários testes de substâncias explosivas no gelo do rio Neva em São Petersburgo.

Avaliação do setor público impregnada de pesquisa é uma adição recente a uma grande cadeia de tentativas por príncipes de usar os recursos intelectuais de estudiosos e cientistas para promover os interesses do Estado. Os serviços solicitados de especialistas de avaliação são, é claro, completamente diferentes dos insinuados nos exemplos citados acima. Estudiosos da avaliação são convidados a

fornecer avaliações dos efeitos de medidas governamentais de conteúdo, administração, *output* e *outcome*, a fim de trazer à autorreflexão uma compreensão mais profunda e decisões bem fundamentadas por parte dos responsáveis das operações do governo. Descartando a noção política banal que boas intenções são suficientes, a avaliação baseia-se na ideia oposta de que as boas práticas e resultados sólidos são o que realmente contam. Avaliação implica olhar para trás, a fim de andar melhor em direção ao futuro. É um mecanismo de vigilância e classificação de atividades do setor público e de seus resultados, de modo que os funcionários públicos, num futuro trabalho, sejam capazes de atuar da forma mais responsável, criativa e eficiente possível.

No entanto, a avaliação sistemática não serve apenas aos príncipes contemporâneos e acadêmicos. Eles são usados pela oposição política, pelas profissões, pelos cidadãos, ou pela clientela dos programas de governo. E eles podem ser realizados por pessoas comuns, não apenas pesquisadores profissionais. Atualmente, a avaliação se tornou uma preocupação não apenas do homem de ciência, mas também do administrador, do profissional, e do homem comum.

1.1 Avaliação Definida

A avaliação é o processo de determinar o mérito, o valor e significância de alguma coisa (SCRIVEN, 1991, p. 1). Implica distinguir o que vale a pena a partir do sem valor, o precioso do inútil.

A avaliação é um procedimento analítico-chave em todos os esforços intelectuais e práticos disciplinados. Embora reconhecendo que o processo de determinar mérito, valor e significância das coisas permeia todos os domínios do pensamento e da prática, na presente avaliação do trabalho, ele será delimitado de acordo com as exigências do serviço público e dos assuntos governamentais. Para o propósito deste livro, propõe-se a seguinte *definição mínima*:

AVALIAÇÃO = df. Aferição cuidadosa *ex post* do mérito, do valor e da significância dos conteúdos, da administração, do *output*, do *outcome* e da organização das intervenções do setor público, que se destina a desempenhar um papel no futuro, em situações de ação prática.

Essa definição de avaliação é controversa. Na verdade, o termo avaliação tem atraído tantos significados diferentes que podemos chamá-lo um ímã semântico (LUNDQUIST, 1976, p. 124). Isso chegou a significar quase qualquer esforço de pensamento sistemático no setor público. É fácil concordar com a primeira frase do livro de Carol Weiss *Evaluation Research* (1972a): “A avaliação é uma palavra elástica, que se estende para cobrir julgamentos de vários tipos”.

Já que avaliação vem em muitas formas, haverá a tentativa de comparar mais detalhadamente outras definições acadêmicas de avaliação com o proposto aqui. Os leitores são aconselhados a ter em mente que neste capítulo só vai ser discutido o significado do termo “avaliação”. Não será dito nada sobre com o que boas ou más avaliações se parecem. Além disso, a definição é mínima, não máxima¹. Qual é o menor conjunto razoável de propriedades que uma investigação no setor público pode ter para que seja nomeada uma “avaliação”?

1.2 Avaliação não é Aferição *ex ante*

Avaliação *não* é aferição *ex ante*. Um exame cuidadoso de intervenções sociais planejadas não é considerado avaliação neste livro. Avaliação é aferição pós-fato, estimação retrospectiva, não aferição antes do fato, estimação prospectiva, consideração premeditada, prognóstico. Avaliação é feita *ex post*, focando em atividades do setor público que estão em curso ou terminadas.

Essa limitação para estimação de intervenções já adotadas em curso ou terminadas é controversa. Grandes teóricos argumentam que avaliação pode também ser aferição *ex ante*, i.e., uma aferição sistemática de *intervenções consideradas, mas não ainda adotadas*. Somando-se às intervenções adotadas em curso ou terminadas, avaliação também concerne a análises preparatórias e *aferições* de várias frentes em intervenções planejadas, mas não ainda adotadas. Para esses teóricos, avaliação é um termo guarda-chuva, cobrindo todos os tipos de análise de, em e para uma intervenção pública.

¹ Referências de definições mínimas são difíceis de se achar. Em Sartori (1984) se encontram argumentos, embora raros, nas páginas 55-56, 64, 70 (“definição mínima”) e 81 (“parcimônia na definição”). Também em Sartori (1976, p. 58 et seq). e Sartori (2009a e b), Collier e Levitsky (2009) e Gerring e Barresi (2009) ver os verbetes Definition, minimal and Parsimony.

Talvez isso seja uma instância de magnetismo semântico da palavra “avaliação”. E não há nada de errado com isso. A palavra pode muito bem ser usada neste sentido amplo. Casos dessa ampla perspectiva de avaliação podem ser observados particularmente no trabalho de *economistas*, que praticam análises de custo-benefício e de custo-eficácia de futuras opções potenciais, mantendo sua avaliação prática.

“Se planejamento é tudo, talvez seja nada?” Aaron Wildavsky (1973) ironicamente perguntou muitos anos atrás na então elegante mania, planejamento. Pode a mesma pergunta ser formulada hoje quando tudo está incluído em avaliação? Claro, é fútil e bobo legislar sobre o uso de uma palavra. Porém se avaliação pode abranger todos os tipos de análises na vida política e administrativa, o conceito não se tornará muito diluído?

O principal argumento contra incluir aferições *ex ante* em avaliação é extraído da emergência da pesquisa avaliativa. As demandas do início do movimento da avaliação por dados empíricos em políticas e resultados de programas se levantaram em oposição à ênfase predominante na análise de intervenções planejadas. Se for permitido que a avaliação abranja até o planejamento, essa significativa linha histórica de conflito será obscurecida.

Assim, neste contexto, avaliação foi limitada às aferições pós-fato no sentido de estimação de políticas já admitidas em curso ou terminadas, programas, componentes do programa, projetos, práticas profissionais, organizações e derivados. É importante notar que pós-facto e *ex post* significam *depois* de a intervenção ser admitida, adotada, estabelecida. *Não* significa depois de a intervenção acabar. Aferições cuidadosas de intervenções em curso são incluídas na definição (avaliação *ex durante*) assim como as intervenções terminadas. Ainda, estimações cuidadosas de ensaios de campo provisórios de políticas e programas também contam como avaliações. No presente livro, excluem-se apenas as análises antes do fato, de intervenções potenciais, consideradas, mas ainda não adotadas.

1.3 Avaliação é uma Aferição Cuidadosa

A avaliação é uma aferição *cuidadosa* de intervenções públicas. Isso significa que a avaliação deve atender a alguns padrões mínimos de qualidade, como a coleta sistemática de dados e aplicação consciente dos critérios de mérito, valor e significância.

Muitos autores célebres foram além e apresentaram o argumento explícito de que a avaliação, por definição, é *pesquisa social*. A seguinte declaração de Eleanor Chelimsky (1985, p. 7) é típica desses estudiosos: “A avaliação do programa é a aplicação de *métodos sistemáticos de pesquisa* para aferição do delineamento do programa, implementação e eficácia (destaque dos autores)”.

É óbvio que Chelimsky está preocupada com os casos em que os cânones da metodologia das ciências sociais são usados para fazer o processo de julgamento mais preciso. Ela tenta traçar uma linha nítida entre pesquisa avaliativa e avaliações improvisadas que confiam na intuição, opinião e sensibilidade treinada. A mesma atitude é demonstrada por Edward Suchman em seu pioneiro *Evaluative Research* (1967, p. 12)²: “Nós não vemos o campo da avaliação como tendo qualquer método diferente do método científico; pesquisa avaliativa é, em primeiro lugar, *pesquisa* e, como tal, deve aderir tanto quanto possível aos padrões de metodologia de pesquisa aceitos atualmente”.

Este livro discorda de Chelimsky, Suchman e outros que prescrevem que a avaliação deve ser minimamente *definida* como pesquisa social, apesar de aqui a pesquisa ser utilizada em um sentido relativamente amplo. Pesquisa não é equiparada à ciência e à aplicação de métodos quantitativos. Ela também inclui inquérito humanista e métodos qualitativos. Além disso, pesquisa abrange não só a investigação fundamental, mas também a comissionada, contratada por terceiros. Todavia, avaliação não será igualada à pesquisa, porque há formas alternativas para tratar de avaliação, não sendo possível apenas descartar por estipulação linguística. Por exemplo, na última década, muitos teóricos de avaliação argumentaram que a avaliação deveria ser “democrática” e realizada por clientes comuns, os cidadãos e as partes interessadas, e não por pesquisadores acadêmicos. Além disso, a avaliação é muitas vezes parte integrante dos processos de tomada de decisão do setor público nos quais procedimentos das ciências sociais desempenham apenas um papel menor. Deve-se também referir estes dois casos à avaliação, desde que haja alguma aferição *cuidadosa* das intervenções envolvidas em curso ou acabadas.

² Veja também Mann (1972); Caro (1971); Meyers (1981, p. 50 et seq.); Nachmias (1979, p.1 et seq.)

1.4 Avaliação concerne às *Intervenções Governamentais*

Uma vez que a avaliação é um processo analítico verdadeiramente geral, ela pode ser aplicada a qualquer área de empreendimento social. No presente livro, no entanto, a avaliação é limitada apenas a *intervenções governamentais*, ou seja, aos empreendimentos, às ações, aos esforços do setor público, como políticas públicas, programas, serviços, projetos e práticas profissionais e organizacionais. Atividades contratadas com as empresas privadas pelos governos estão naturalmente inclusas.

Intervenções públicas contemporâneas cobrem tanto programas substantivos, quanto orientados aos processos (LUNDQUIST, 1990). Medidas substantivas se referem a diversos domínios funcionais tais como energia, meio ambiente, recursos naturais, uso do solo, habitação, assistência social, saúde, transporte, fomento ao desenvolvimento e muitos outros campos.

Intervenções orientadas aos processos — reforma administrativa — referem-se às ideias e medidas dirigidas à organização e funcionamento da própria administração pública. A reforma administrativa está mais preocupada com a gestão por objetivos do que gestão orientada aos processos, à descentralização, aos novos sistemas de orçamento, às mudanças na administração local e a outros processos de desenvolvimento institucional. Um problema central na reorganização administrativa contemporânea diz respeito a quais são os arranjos institucionais que devem ser usados na implementação de intervenções públicas: agências reguladoras repletas de funcionários executivos neutros e competentes, pessoal nomeado pelos seus méritos políticos, execução através dos municípios, arranjos corporativos, parcerias locais público-privadas, agrupamentos de empresas privadas, departamentos universitários e órgãos públicos, profissionais, envolvimento do cliente ou subcontratação de empresas privadas (quadro geral em LUNDQUIST, 1985).

Não é preciso dizer que a avaliação engloba a aferição de intervenções substantivas, bem como as intervenções governamentais orientadas aos processos. A avaliação é dirigida a todos os tipos de atividades do setor público.

1.5 Avaliação é Focada em *Conteúdo da Decisão, Administração, Outputs e Outcomes*

Como definido aqui, a avaliação não está preocupada com todo o ciclo de políticas, mas apenas com a parte final do mesmo. Para deixar claro, será introduzido o pensamento sistêmico, tão predominante no estudo da administração pública com o prisma da ciência política. Um *sistema* é um todo com componentes que são dependentes entre si. Na sua forma mais rudimentar, um sistema consiste de *input*, conversão, e *output*, como na Figura 1.1.

Figura 1.1 – O Modelo Sistêmico Simples



O modelo sistêmico simples é extremamente geral e abstrato. Ele é usado em muitas ciências. Em biologia, por exemplo, o ser humano pode ser considerado como um sistema, com *input* (alimentos, água), conversão (digestão) e *output* (trabalho produtivo, resíduos).

Isso também é utilizado em políticas públicas e administração pública. A agência do governo pode ser vista como um sistema, como pode uma escola, ou uma comissão municipal setorial para o meio-ambiente. Mas, quando se aplica o modelo sistêmico simples aos fenômenos do setor público, algumas funções são adicionadas e a terminologia se altera um pouco. Uma agência federal de governo pode ilustrar a ideia geral. O *input* para uma agência proveniente do seu principal, por exemplo, do governo, podem ser fundos com algumas condições para seu uso, instruções escritas, apoio ou crítica oral, e pessoas designadas. Dentro da agência, os fundos, as pessoas e as instruções são coordenadas e convertidas em outra coisa. A conversão é o que está acontecendo no interior da agência. O *output* é o que resulta da agência.

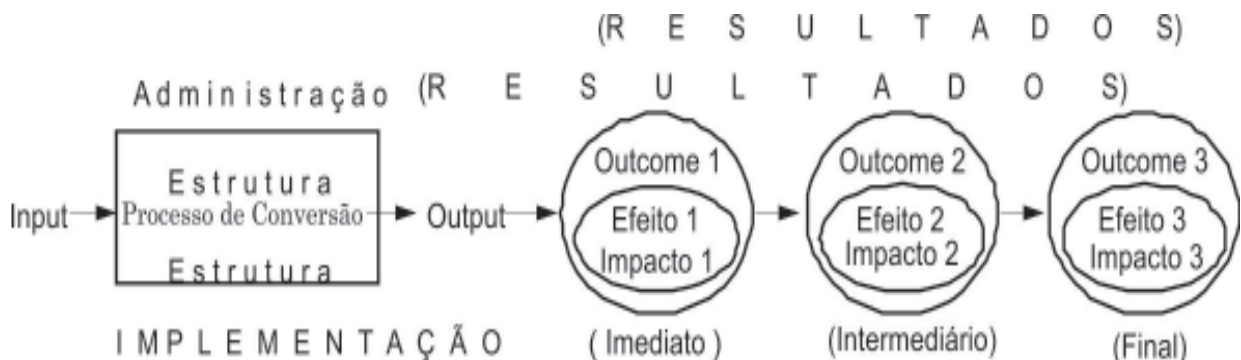
Nos estudos de políticas públicas, o estágio de conversão do modelo geral de sistemas é aproximadamente equivalente à administração e uma fase de *outcomes* muitas vezes é fixada na fase de *output* do modelo de sistema geral. Entende-se por *output*

fenômenos que resultam dos órgãos do governo na forma de, por exemplo, proibições, permissões de procedimentos, subvenções, subsídios, impostos, exortação, indução, persuasão moral, serviços e bens.

Outcomes ocorrem quando os *outputs* chegam aos destinatários, incluindo as ações dos destinatários, mas também o que acontece além dos destinatários na cadeia de influência. É possível distinguir *outcomes* entre imediatos, intermediários e finais. Outro termo para *outcome* é impacto.

'Resultados' será usado como termo tanto para *outputs*, como para *outcomes* ou *outputs* e *outcomes* juntos. O termo 'implementação' usualmente cobre conversão e *output*, ou somente conversão ou ambas, mas não *outcome*. O raciocínio está resumido na Figura 1.2.

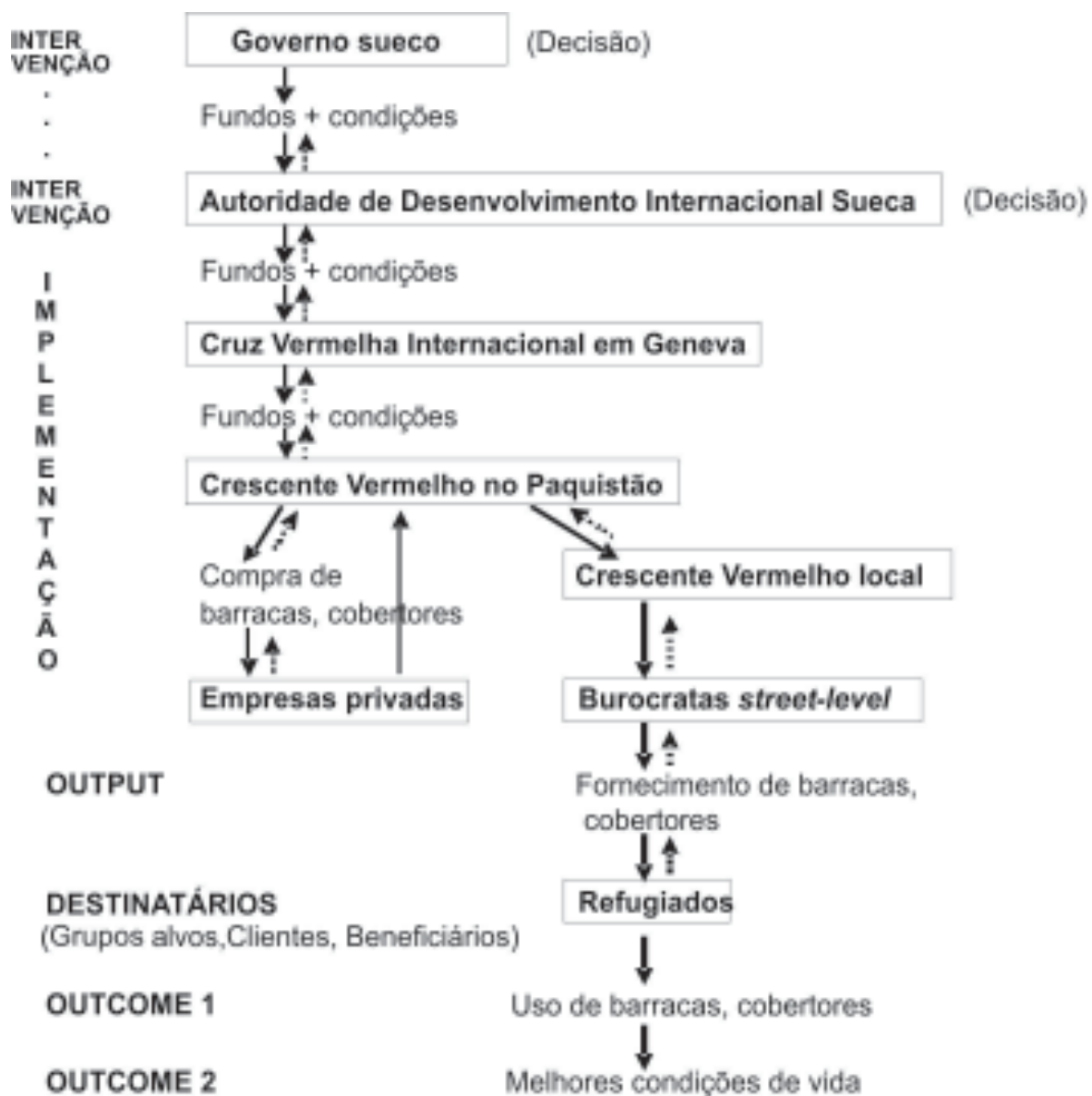
FIGURA 1.2 - O Modelo Sistêmico Adaptado para Avaliação de Intervenção Governamental



Ilustrando com um exemplo meio-realista, meio-imaginário, como na Figura 1.3. Alguns anos atrás, autoridades suecas instituíram um programa para ajudar os refugiados da guerra civil no Afeganistão, que viviam em campos no Paquistão. O governo e a Agência Sueca de Cooperação para o Desenvolvimento Internacional, abreviada SIDA (em inglês), alocaram recursos e concordaram com o objetivo de proporcionar tendas e cobertores aos refugiados. Para implementar esse programa, SIDA fechou um acordo com a Cruz Vermelha Internacional, em Genebra, que prometeu canalizar o dinheiro, sob certas condições, para o Crescente Vermelho Nacional no Paquistão.

O Crescente Vermelho Nacional foi instruído a utilizar os fundos para comprar barracas e cobertores de comerciantes locais privados, distribuí-los para o Crescente Vermelho local pertinente e seus burocratas de nível de rua (*street-level*³), que deveriam designar barracas e cobertores para os refugiados nos acampamentos. Como resultado do uso do equipamento, os refugiados teriam seu sofrimento aliviado.

Figura 1.3 - O Programa de Ajuda para Refugiados Afegãos



³"Street-level bureaucrats" são operadores que têm contato direto com o alvo da intervenção na linha de frente. Em português, burocratas de nível de rua.

Para se qualificar como uma avaliação, um estudo do Programa de Ajuda para Refugiados Afegãos adotado deve se concentrar seja: (1) nos *outcomes* – se os refugiados estavam realmente usando as tendas e os cobertores e se isso criou melhores condições de vida para eles; (2) no *output* – o fornecimento de tendas e cobertores através dos burocratas *street-level* do Crescente Vermelho paquistanês local; (3) na administração – o que aconteceu com os fundos e os objetivos, uma vez que a SIDA iniciou as negociações com a Cruz Vermelha Internacional por meio do Crescente Vermelho paquistanês e com as empresas privadas desde o início da distribuição de barracas e cobertores pelos burocratas *street-level*, ou, finalmente; (4) no conteúdo da intervenção adotada – foram as metas estabelecidas e os meios planejados potencialmente relevantes e adequados para aliviar os problemas sociais definidos na decisão do programa?

Reconhecidamente, a avaliação dos *outcomes* pode ser considerada mais importante do que as avaliações de *output*, conversão e avaliação do conteúdo da intervenção. No entanto, não se quer aqui equiparar avaliação com avaliação de *outcomes*. O conceito, tal como definido aqui, inclui a preocupação com a relevância e importância do objetivo, com os processos administrativos e com os *outputs* também. No caso dos refugiados, por exemplo, todos podem ver que a administração é um longo curso de ação, onde muitas coisas podem dar errado.

Um esclarecimento adicional é, provavelmente, necessário. A limitação da avaliação como *aferição* de metas, meios e outros conteúdos de intervenção, processos administrativos, *outputs* e *outcomes* não está preocupada com fatores explicativos em avaliação. Se a avaliação se propõe a explicar o que influenciou as variações nos procedimentos administrativos, *outputs* e *outcomes*, a definição aqui apresentada permite que fatores explicativos sejam extraídos de qualquer lugar. Seria vergonhosamente inadequado delimitar o conceito de avaliação no que diz respeito aos determinantes que podem ser observados.

Até agora, o livro se aventurou para justificar a delimitação do objeto de avaliação como “conteúdo, administração, *outputs* e os efeitos dos *outcomes* das intervenções governamentais”. Na próxima seção, será abordado o que significa para a avaliação se preocupar com “atividades em curso e acabadas”.

1.6 Avaliação é Aferição de Atividades em Curso e Acabadas

Às vezes, avaliação é restrita a atividades *em curso*, não incluindo a aferição de políticas e programas finalizados. Essa perspectiva restrita é claramente discernível no livro de David Nachmias, *Public Policy Evaluation* (1979, p. 3 et seq.):

Um método que pode reduzir o número de decisões errôneas é a abordagem formal científica do conhecimento... Vista de uma perspectiva científica, pesquisa avaliativa de políticas públicas é o exame objetivo, sistemático e empírico dos efeitos que políticas *em curso* e programas públicos têm em seus alvos em termos de metas que eles pretendem atingir. (ênfase dos autores)

Na verdade, a perspectiva estreita no assunto da avaliação era um lugar comum na literatura americana e canadense mais antiga. De acordo com Rutman, "avaliação de programa se refere ao uso de métodos de pesquisa para medir a eficácia de programas *vigentes*," (1980, p.17, *itálico dos autores*). Wholey *et al.* escreveu: "Avaliação afere a eficácia de um programa *em curso* quanto ao sucesso em atingir seus objetivos" (1970, *itálico dos autores*).

Indubitavelmente, intervenções em andamento claramente fazem parte dos assuntos relacionados com avaliação. Elas podem até constituir-se no assunto principal da avaliação do setor público (avaliação *ex durante*); mas deveria a avaliação estar confinada *somente* a atividades em curso? A resposta é não. Lições podem ser aprendidas também a partir de aferições de operações acabadas. Por que excluir estas da pesquisa avaliativa? Aferições de atividades públicas finalizadas (avaliação *ex post*) podem muito bem ajudar na melhoria das operações em curso ou planejadas. Delimitar a avaliação somente para o estudo das atividades em curso acabaria por deixar de fora, indevidamente, partes importantes da avaliação do setor público. A avaliação deve incluir todos os estudos retrospectivos das políticas, programas, mudanças organizacionais e afins em curso, bem como das terminadas. De acordo com a definição, avaliação compreende todos os tipos de análises de conteúdo pós-fato, processos administrativos, *outputs* e *outcomes* das atividades do setor público.

1.7 Avaliação é Mais do que Aferição de Impactos Planejados

Metodologistas eminentes também definiram avaliação como análise do efeito planejado (*outcome*). Em seu *Planning Useful Evaluations*, Leonard Rutman (1980, p.17) afirma: “Neste livro, avaliação de programa refere-se ao uso de métodos de pesquisa para medir a *eficácia* de programas operantes”. E eficácia, ele complementa, é até onde se estende o sucesso do programa com relação ao alcance de seus [*outcomes*] objetivos.

Essa estreita delimitação não está errada, é claro. Contudo, não se pode aceitá-la aqui. Avaliação pode ser a análise dos efeitos planejados; porém se acredita, neste livro, que avaliadores devem também traçar efeitos colaterais inesperados e não intencionais nos estágios de *outcomes*. Avaliação deve também incluir análise passo-a-passo da implementação e do *output* a fim de localizar práticas inovadoras assim como barreiras e obstáculos e o que faz com que eles ocorram. Além disso, aferição da eficiência e mensuração simples do alcance de metas, onde a questão dos efeitos não é levantada, também pertencem à avaliação. Ademais, na avaliação da pesquisa acadêmica, bibliotecas, museus, saúde pública, parques públicos e outros serviços do governo, a qualidade do *output* é avaliada, por exemplo, por painéis de peritos, um arranjo de pares, ou grupos de clientes; mas nestes estudos raramente se levanta a questão dos efeitos do *outcome*. Mesmo assim, os estudos são referidos como avaliação, e com razão. Esses estudos não devem ser excluídos da família da avaliação por decreto.

1.8 Metas das Intervenções não são os Únicos Critérios de Valor Permitidos

No discurso da avaliação, não há sequer um acordo num ponto aparentemente fundamental: que avaliação está preocupada com a determinação de mérito, valor e significância. Charles Atkin, por exemplo, não inclui referência a valorização em sua definição, que é a seguinte: “avaliação é a coleção e apresentação de dados resumidos para tomadores de decisões” (citação de SCRIVEN, 1991, p.156). “Só dê a eles os fatos”, essa é a atitude de Atkin quanto à avaliação.

Adota-se, neste livro, uma posição diferente. Avaliação significa valorizar. Por definição, avaliação envolve a aferição de mérito, valor e significância. O componente de valor da avaliação pressupõe pelo menos um critério de mérito a partir do qual as intervenções públicas são julgadas.

Reconhecidamente, a formulação tradicional da questão de avaliação de programa é: até que ponto o programa tem sucesso em alcançar suas metas? (WEISS, 1972a, p. 4; NACHMIAS, 1979, p. 3 et seq.). Mas por que deveriam ser as metas das intervenções os *únicos* critérios permitidos para julgar o valor das políticas públicas? Por que não podem ser utilizados julgamentos usando preocupações e expectativas de clientes, ou de qualquer outra audiência de interessados (*stakeholders*), e isso também ser chamado de avaliação? Excluir todos os parâmetros com exceção das declaradas metas das intervenções é circunscrever o conceito de modo muito restrito. Serão por decreto impossibilitadas discussões com respeito a quais critérios serão aplicados ao que está sendo avaliado. Portanto, a definição aqui proposta afirma que avaliação engendra estimação, mas deixa em aberto que tipos de critérios de valor podem ser usados. Podem ser metas das intervenções; mas as expectativas dos investidores interessados, os desejos dos clientes, critérios profissionais de mérito, e assim por diante, também podem ser usados.

1.9 Critérios de Valor e Mérito Podem não ser Necessariamente Apenas *ex ante*

Avaliação implica na aplicação de critério de valor *antes* da mesma. Este argumento engendra uma desnecessária e pouco frutífera demarcação. Por que a avaliação da educação superior não pode, por exemplo, proceder de um critério de valor que emergiu *durante* o processo de avaliação em diálogos entre profissionais experientes? E por que não pode a aplicação minuciosa *ex post* por parte dos tomadores de decisões de seus próprios critérios de valor aos dados, proveniente de um estudo cuidadosamente executado, ser admitida como avaliação? Além disso, aferições cuidadosas usando critérios que emergem durante um processo de avaliação ou depois que o processo termina são perfeitos exemplos de avaliação.

1.10 Avaliação Deve Ser Útil

Avaliação não é qualquer estimativa cuidadosa; é uma estimativa que pretende fazer parte de situações futuras de ação prática. Para uma estimativa ser chamada de uma avaliação seu propósito tem que ser útil. E ser útil é o suficiente. Não é necessário que uma estimativa seja utilizada para ser chamada de avaliação. Em suma, se a intenção de utilidade não se faz presente, então não é uma avaliação. Essa orientação geral prática de utilidade é tão central que deve ser incluída na definição.

Por outro lado, uma maior especificação da orientação de utilidade faria a avaliação muito circunscrita. Alguns veteranos de pesquisa avaliativa argumentam, por exemplo, que avaliação sempre mira em refinar a intervenção. Lee Cronbach *et al.* (1980, p. 14), entre outros, definiu avaliação com respeito à ideia de melhoramento da intervenção:

Por *avaliação*, queremos dizer exame sistemático de eventos ocorrendo dentro e conseqüentemente num programa contemporâneo — um exame conduzido para ajudar na melhora desse e de outros programas que têm o mesmo propósito geral.

Na verdade, aperfeiçoamento futuro de intervenções é um importante e compreensivo propósito de avaliação. O problema com a *definição* de Cronbach é que numerosos estudos avaliativos encarregam-se de propósitos práticos que não o aperfeiçoamento. Exercer avaliação por motivos de prestação de contas, isto é, os *principais* que empregam a avaliação com o objetivo de manter seus executivos responsáveis pelo o que eles têm feito, é um propósito perfeitamente válido para avaliação (ver capítulo 7). Ademais, avaliações são empreendidas por motivos políticos, por exemplo, para poupar tempo ou distrair interesse público de um assunto. A fim de não impedir longo alcance e escrutínio do porquê avaliação deve ser exercida, parece preferível não limitar avaliação somente ao propósito de melhoramento da intervenção. Por definição, avaliação é um empreendimento que pretende fazer parte de situações futuras de ação prática como, por exemplo, num debate entusiasmado de política

pública ou mais diretamente provendo materiais úteis para alguma decisão pendente. Contudo, o conceito não deverá ser mais delimitado com relação a seu propósito.

1.11 Modelos Sistêmicos são Dispositivos Heurísticos

Finalmente, deve ser ressaltado que os modelos sistêmicos nas figuras 1.2. e 1.3. *não devem* ser entendidos como representações de um mundo existente. Não são modelos empíricos, indicando que *inputs* são transformados diretamente em *outputs* que seriam por sua vez transformados em *outcomes* ou efeitos dos *outcomes*. Nem expressam ideais normativos de como *inputs* devem ser transformados em *outputs*, e *outputs* em efeitos do *outcomes*, no melhor dos mundos.

Os modelos sistêmicos são *ferramentas heurísticas* que ajudam os investigadores a levantar questões, a mapear processos reais, ou a penetrar argumentos normativos. “Heurístico” é derivado da palavra grega para “achar”. Encontrei! Descobri! “*Eureka!*”, a exclamação de Arquimedes em seu banho depois de o que é conhecido hoje por princípio de Arquimedes (“um corpo imerso em um líquido sofre um empuxo para cima por uma força igual ao peso volume do líquido deslocado pelo corpo”). Uma ferramenta heurística é consequentemente um instrumento de trabalho através do qual podemos levantar questões interessantes para descobrir como o mundo realmente funciona ou para trazer ordem para uma controvérsia normativa. Críticas dos modelos sistêmicos nas figuras 1.2 e 1.3, concebendo que o mundo não é tão parcimonioso e bem organizado como os modelos indicam ou que os modelos representam uma utopia inatingível, são, então, inúteis. Relevantes são, por sua vez, as objeções, mostrando que os modelos geram questões pouco interessantes ou desconsideram os problemas centrais.

Isso encerra a justificação para definir avaliação como uma aferição cuidadosa *ex post* do mérito, valor e significância do conteúdo, da administração, do *output* e efeitos dos *outcomes* e da organização de intervenções governamentais, que se destinam a desempenhar um papel nas futuras situações de ação prática.

INTERVENÇÃO, AVALIAÇÃO, FEEDBACK, META-AVALIAÇÃO

Avaliação contém um componente retrospectivo e um prospectivo. Ela nos leva a olhar para trás – o componente retrospectivo – para melhor orientar o que vem pela frente – o componente prospectivo. O papel da avaliação é o de acumular e avaliar sistematicamente informações sobre o conteúdo das intervenções, a sua administração, *outputs*, os efeitos dos *outcomes*, e organização – baseada em fatos passados – com a finalidade de produzir informações úteis que podem conduzir a ajustes, ou, em geral, decisões futuras mais racionais.

2.1 O Modelo Multicamada de Governança

Uma forma superior de ilustrar a ideia de uma consideração retrospectiva orientada prospectivamente é ver o processo de intervenção e o processo de avaliação como dois, como atividades em camadas, onde o último, processo de segunda ordem, reflete sobre o processo anterior, de primeira ordem, com o fim de exercer alguma influência sobre ele.¹ Uma tarefa crucial nos processos de governança é ligar este processo de segunda ordem ao processo de intervenção de primeira ordem. Isto se consegue através de um mecanismo de *feedback* em que o processo de avaliação e os resultados são utilizados. Além disso, o *feedback* pode ser percebido separadamente, como se fosse um processo de terceira ordem. No Modelo Multicamada de Governança apresentado na figura 2.1, o processo de intervenção é descrito como um processo de primeira ordem, com avaliação e *feedback* desenhados como dois meta-processos distintos de segunda ordem e terceira ordem.

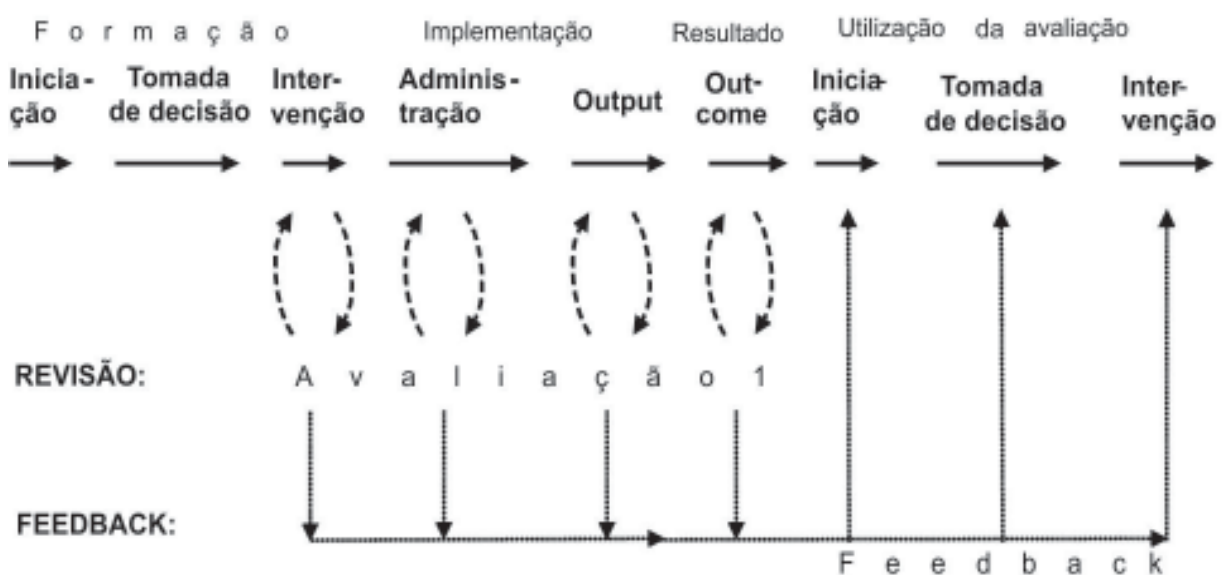
¹ A ideia de tratar avaliação como um meta-processo é claramente exposta em Fernández-Ballesteros (1992a, p. 207) e (1992b), que distingue o ciclo de intervenção de política pública de funções de avaliação (aferição). O chamado modelo de avaliação da União Europeia de avaliação no capítulo 4 inclui uma exposição bastante clara desse tema.

A avaliação não é o único mecanismo público que supostamente garante esta reflexividade. Ao contrário, num Modelo Multicamada de Governança avaliação é considerada como parte de um processo de revisão abrangente. A revisão é um grande guarda-chuva como método para garantir que os processos de intervenção ocorram sem problemas e de modo sensato. Mecanismos de revisão também incluem inspeção, monitoramento, auditoria financeira, e de recurso.

É importante salientar mais uma vez que o modelo de governança na figura 2.1 *não* oferece uma descrição empírica de qualquer realidade. É um instrumento *heurístico* bruto, um instrumento idealizado para ajudar o pensamento. Ele deve orientar a reflexão e a pesquisa sobre processos de intervenção pública, avaliação e utilização da avaliação.

Figura 2.1 - Avaliação no Modelo Multicamada de Governança

PROCESSOS DE INTERVENÇÃO:



Na terminologia geral de sistemas, representada no capítulo anterior, a iniciação, a tomada de decisão, e a intervenção no processo intervencionista pertenceriam à função *input*, a administração seria a conversão, e o *output* seria o *output*. Os *outcomes* normalmente não são cobertos no modelo geral de sistemas, mas devem ser incluídos em qualquer modelo heurístico com intuito de ajudar os processos de avaliação.

No Modelo Multicamada de Governança apresentado na figura 2.1, as funções do processo de intervenção pública são limitadas a seis. Ela começa com a formação, que consiste, por sua vez, da iniciação, tomada de decisão e intervenção. *Iniciação* diz respeito à identificação de um problema, cuja solução supostamente exige ação pública. Tais demandas podem surgir de dentro do sistema político, por exemplo, por um órgão administrativo, ou a partir de fontes externas, tais como organizações de interesse ou a mídia. Quando autoridades políticas identificam o problema, ele entra na agenda política. Quando os políticos nacionais se preocupam e tentam levá-lo para algum canal da agenda política, então o problema é divulgado na agenda política nacional. A iniciação pode ocorrer de forma abrupta e rapidamente, por exemplo, em crises repentinas.

A *função de tomada de decisão* consiste na preparação de decisões e adoção de intervenção autorizativa.² Na *preparação de decisões*, são feitos esforços para identificar a natureza do problema com o intuito de criar as bases para uma futura decisão. Uma parte será sobre o tamanho do problema no presente, como ele se desenvolveu ao longo dos anos, e como ele provavelmente irá se desenvolver se nada for feito. As causas do problema serão outro tema. Um terceiro tema são, geralmente, as consequências do problema no presente, no passado e no futuro se nada for realizado para resolver. Então haverá seções sobre metas a serem estipuladas e alcançadas e os instrumentos de políticas públicas que poderão ser elaborados e como eles afetarão o problema e as suas consequências.

A preparação de decisão pode envolver inúmeros atores, procedimentos e práticas, por exemplo, a nomeação de um órgão formal de investigação, apresentação de opções preliminares para ação, a revisão ou análise pelos grupos de interesse e por outras partes interessadas. Essa preparação usualmente termina com a apresentação de uma ou mais recomendações finais para ação junto com o resto do material informativo.

² “Decisão” em “tomada de decisão” ou “processo decisório” e “preparação da decisão” significa decisão de intervenção. Evidentemente numerosas decisões são também feitas em outros estágios do processo de intervenção. Decisões são feitas em relação a como a implementação deve ser executada, como o controle *ex-post-fato* deve ser organizado, e assim por diante. Esta óbvia complicação será negligenciada aqui.

A preparação de decisão é seguida pela *adoção final de uma intervenção autorizativa*. Quem vai fazer esta resolução formal, legitimadora varia, obviamente, de intervenção para intervenção. Se a adoção diz respeito à legislação, ela deve ser feita por um parlamento. Além disso, decisões de intervenção podem ser feitas pelo governo, por uma agência nacional e/ou por uma agência regional. No nível municipal, as decisões são tomadas pelos conselhos, câmaras de vereadores, comissões ou agências municipais. Em uma escola, por exemplo, as decisões são tomadas pelo conselho, pelo diretor, pelo reitor ou por um grupo de professores.

A adoção de uma intervenção refere-se a uma decisão formal de intervenção. O lugar específico onde as decisões reais são feitas é mantido aberto para uma investigação empírica.

A *intervenção* neste caso representa o conteúdo substantivo da intervenção formalmente aprovada. Uma intervenção geralmente contém seções sobre problemas subjacentes, objetivos e instrumentos de políticas públicas.

A próxima parte do modelo multicamadas cobre a implementação, dividida em administração e *output*. *Administração* se refere às atividades principais uma vez que a adoção da intervenção é transformada, mas não incluindo em *output*. A ideia básica é que a intervenção adotada deve ser levada adiante para a sua realização. Através de decisões administrativas, esforços são feitos para planejar, projetar e ter as intervenções prontas para ser entregues aos alvos. Este é o processo principal da administração. No entanto, o processo principal precisa de processos de suporte como contratação e treinamento de pessoal, aquisição de equipamento técnico, pagamentos de salários e de comunicação com o mundo exterior.

Administração inclui coisas como a difusão pelos tomadores de decisão de nível superior de esclarecimentos sobre o conteúdo da intervenção, e a emissão por parte dos governos e agências reguladoras de normas detalhadas e regras sob os regimes regulatórios gerais (como, por exemplo, a Declaração de Combustível no capítulo 3). A preparação para a execução efetiva refere-se a tentativas políticas para afetar órgãos do governo, as tentativas da alta gerência em influenciar a gestão inferior, e as tentativas das gerências de menor escalão em influenciar os operadores *street-level*. Os destinatários também podem contribuir com a administração, por exemplo, através da participação ativa na formação da intervenção feita por uma agência de base. Na linguagem da administração pública, isso é chamado de participação do cliente ou do usuário.

Nos setores públicos de hoje em dia, a administração é realizada por vários outros atores que não as agências formais, como as empresas comerciais, organizações de interesse público, organizações de interesses, e várias formas de parcerias público-privada (consulte o caso do Afeganistão no capítulo 1 e o caso da Declaração de Combustível no capítulo 3).

Os *outputs* são tudo em um processo de intervenção que, em última análise, podem influenciar os alvos de intervenção. Normalmente se tem em mente esses meios através dos quais os operadores *street-level* tentam influenciar alvos de intervenção. *Outputs* são as coisas que estão realmente transbordando do nível mais baixo de um sistema administrativo. Eles incluem serviços sociais e bens entregues, leis e regulamentos aplicados a casos individuais, impostos e taxas cobradas, subvenções e empréstimos desembolsados, mensagens das campanhas publicitárias distribuídas aos alvos, e assim por diante. Em termos mais concretos, o que acontece é que os professores ensinam nas escolas, os enfermeiros assistem idosos nos abrigos municipais, e os leitores tomam livros emprestados das bibliotecas públicas.

No entanto o conceito de *output* é mais complicado que isso. A preparação da decisão bem antes de uma intervenção ser formalmente adotada pode estimular e influenciar alguns alvos a agir bem antes do que o esperado. O mecanismo é a *antecipação*. Neste caso, os eventos ocorridos durante a preparação de decisão são na realidade *outputs*, e os efeitos dos resultados são produzidos antes de a intervenção ser formalmente inaugurada (VEDUNG; SVÄRD, 2008; e VEDUNG, 1993, p.216 et seq., p.221, p.241, p.265 e p.270 et seq.). Ainda mais difícil de entender, talvez, é onde a intervenção adotada pode funcionar como um *output* aos olhos do destinatário. Nesta situação, a intervenção em si pode influenciar os destinatários a obedecer ao que foi prescrito na ação, e não às atividades de implementação por alguma agência que procura ser a executora. Consequentemente, o *output* não é somente aquilo que sai do sistema de execução no nível mais básico.

No Modelo Multicamada de Governança, *outcome* é colocado como uma função separada depois do *output*. Este é, talvez, o seu desvio mais óbvio a partir do modelo sistêmico simples. *Outcomes* são eventos, fatos que acontecem com os destinatários (beneficiários, clientes, alvos) junto com o processo de intervenção. Em algumas circunstâncias, *outcomes* são descritos e avaliados como as atitudes, ações e desejos de agir dos destinatários. Em outras situações, os *outcomes* são essencialmente constituídos por repercussões sociais

primárias, secundárias, ou até mesmo terciárias para além dos destinatários. O verdadeiro desafio é destacar as partes dos *outcomes* que, pelo menos até certo ponto, são efeitos da intervenção pública e sua implementação (= *outcomes* líquidos, efeitos líquidos) de *outcomes* em geral (= *outcomes* brutos).

Em todas as organizações, a tarefa fundamental da gestão é a *revisão*, depois do fato, para fins de controle. Um dispositivo é necessário para o monitoramento, redirecionamento, convencimento e desenvolvimento das diversas e especializadas atividades das grandes organizações complexas, tais como uma burocracia pública, se o sistema pretende ser efetivo, eficiente e legítimo. No presente contexto, a revisão depois do fato com fins de controle é vista como um meta-processo, um ciclo separado, contínuo de auditoria financeira tradicional, bem como de monitoramento do processo de intervenção pública primária, incluindo a coleta de informações sobre o desempenho de seus subordinados, e o rastreio do cumprimento pelos destinatários. A revisão também gera inspeção na forma de visitas, seguido por atividades como consultoria, diálogo e intimações.

A revisão também inclui avaliação. Os tomadores de decisão, a gestão superior, a gestão intermediária e os operadores *street-level* autorizam os seus subordinados ou pesquisadores externos e consultores para descobrir como as várias etapas do processo de administração estão se desenvolvendo, como os *outputs* e *outcomes* se mostram, se os *outcomes* são produzidos pela intervenção, e se existem meios mais eficientes em termos de custo para alcançar o mesmo objetivo. A avaliação pode também ser conduzida como autoavaliação com a finalidade de aprendizagem e de autoajuste.

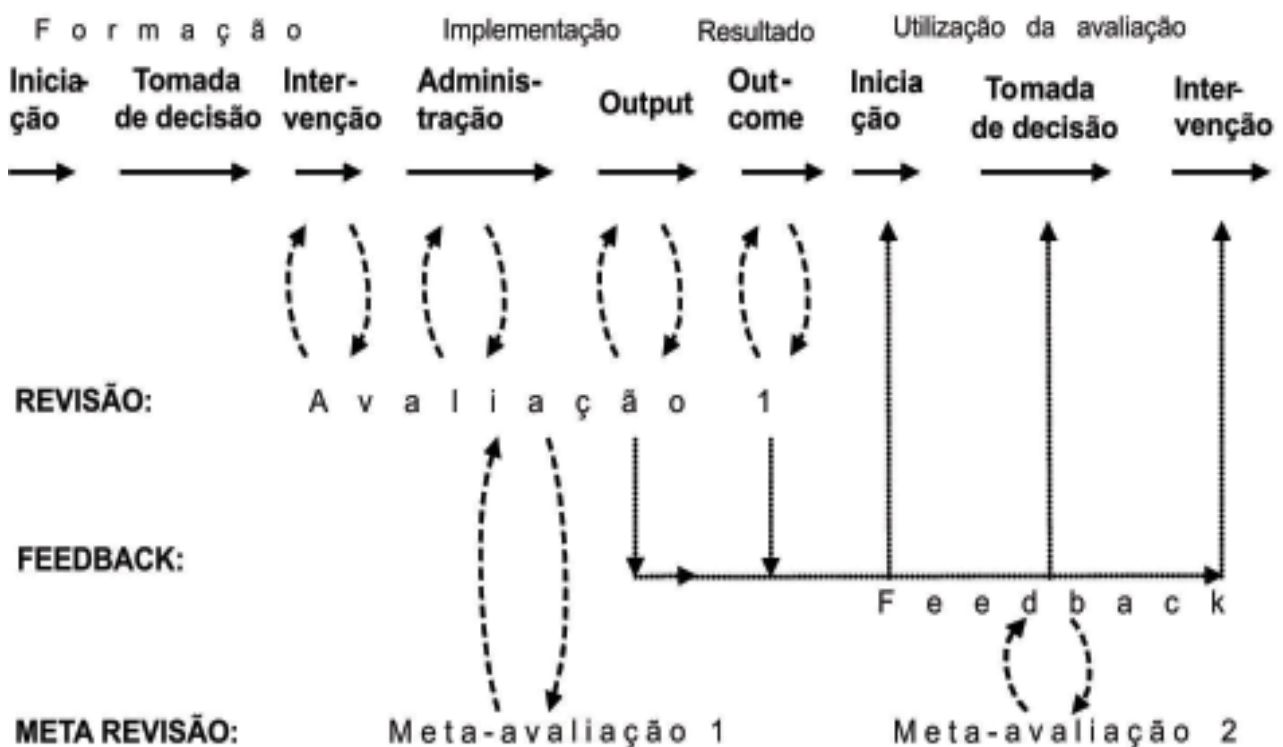
Resultados descobertos através da revisão, em particular por meio da avaliação e inspeção, são realimentados aos atores do processo de intervenção. A entrega das informações da revisão é chamada de *feedback* na figura 2.1. *Feedback* deve produzir reforma, redirecionamento e reconsideração. Neste estágio de redirecionamento e reconsideração, três coisas podem acontecer como uma consequência do aprendizado: continuar como antes, mudança ou término. Com base nas novas circunstâncias, o processo de intervenção deve ter mais uma rodada e ser submetido a um controle pós-fato, o que dá início a um novo *feedback*, etc.

2.2 Meta-avaliação e Auditoria

A quarta camada do Modelo Multicamada de Governança é a meta-revisão na forma de meta-avaliação 1 (avaliação das avaliações, avaliação da função de avaliação) e meta-avaliação 2 (avaliação de feedback de avaliação(ões))³. Meta-avaliação baseia-se em percepções de que os processos de segunda e terceira ordem, como avaliação e *feedback* são também intervenções que podem ser avaliados. Avaliação e o *feedback* da avaliação como partes da gestão em uma organização, são avaliáveis e auditáveis como outros processos de gestão, como implementação de políticas. As relações entre os variados níveis estão ilustradas na figura 2.2.

Figura 2.2 - Meta revisão como Meta-avaliação 1 e 2 no Modelo Multicamada de Governança

PROCESSOS DE INTERVENÇÃO:



³ *Quis custodiet ipsos custodes?* Quem guarda os guardas? Quem vigia os vigias? Quem olha a sentinela? Quem policia a polícia?

Ademais, conhecimentos produzidos pela avaliação da avaliação e de avaliação do *feedback* da avaliação de devem ser mandados de volta para os tomadores de decisões apropriados, a fim de serem utilizados. Esse mecanismo de meta-*feedback* não está ilustrado na figura 2.2.

Uma questão frequente na literatura de avaliação (HOOGERWERF, 1992, p. 215 et seq.) é a de que a meta-avaliação é especialmente apropriada quando os superiores desejam avaliar o desempenho dos órgãos subordinados com forte base científica-profissional. A pesquisa universitária é caso perfeito em que o governo dá apoio a atividades onde os profissionais reinam quase que supremamente. A ideia é deixar os profissionais comandarem a função da avaliação por eles mesmos – professores são requisitados para avaliar outros professores e seu desempenho, pois profissionais de fora não teriam a competência adequada – mas de modo que um tomador de decisão superior possa determinar se os mecanismos de avaliação estão de fato funcionando, em outras palavras, praticando a meta-avaliação. Pode parecer difícil de acreditar, mas um autor sugeriu meta-meta-controle (sic!) como um meio apropriado de auditoria para que os sistemas de auditoria não percam o gás ou saiam da linha porque as entidades auditadas poderiam enganar os auditores (POWER, 1997, p.144 et seq.).

A ideia da meta-avaliação mostra que avaliação deve ser um sistema autorreferencial. Avaliação reflete em outras intervenções, mas é uma intervenção por si só que pode refletir nela mesma.

É necessário enfatizar mais uma vez que o Modelo Multicamada de Governança é um modelo heurístico, que supostamente facilita o pensamento sobre a avaliação. Tem o propósito de ajudar avaliadores e comissários de avaliação a colocar-se num contexto mais amplo. Ela não fornece nem pretende fornecer uma visão realista do que está acontecendo nos governos. A imagem realista pode ser pintada apenas com base em cuidadosos estudos empíricos.

TEORIAS DE INTERVENÇÃO COMO INSTRUMENTOS DE AVALIAÇÃO

Um programa pode ser concebido como uma hipótese:
se um programa é implementado,
então o resultado esperado terá êxito.

**John McLaughlin A & B Gretchen Jordan,
"Using Logic Models", (2004, p. 8).**

O método da teoria de intervenção é uma ferramenta valiosa para avaliação. O pensamento em termos de teorias de intervenção (teorias de programa) é desenvolvido dentro do campo de estudos de avaliação. Esta não é a única contribuição da avaliação para o pensamento social, mas, provavelmente, a mais importante. Na Ciência Política, essa ferramenta ainda não ganhou um apoio, embora os cientistas políticos gastem uma grande quantidade de esforço em estudos de políticas públicas e análise de ideias políticas. Os praticantes da ciência política, administração de empresas, serviço social, saúde pública, geografia cultural, planejamento da comunidade, e outras disciplinas ainda têm muito a aprender com a avaliação a este respeito. No entanto, a teoria de intervenção também tem desvantagens, mais significativamente a inobservância de efeitos colaterais inesperados.

Neste capítulo, o leitor vai conhecer o que se entende como uma teoria de intervenção, como tais teorias são interpretadas e estruturadas, a fim de que se tornem instrumentos de avaliação adequados, e como esses instrumentos podem ser utilizados. As duas partes principais de uma teoria de intervenção e os pontos fortes dessa abordagem serão discutidos. A inobservância de efeitos surpreendentes e outras fraquezas também serão examinadas.

3.1 O Que é uma Teoria de Intervenção?

A palavra "teoria" na expressão "teoria de intervenção" desperta curiosidade e espanto. Geralmente, o termo é utilizado como um

antônimo de “prática”. Na maioria das vezes, a teoria de intervenção refere-se a crenças sobre como a intervenção deveria funcionar, e não como ela realmente se desenrola no mundo real. Para evitar falsas associações, o termo “lógica da intervenção” (a lógica do programa) e equivalentes estão ganhando terreno.

3.2 Inserido na Intervenção Institucionalizada

Em matéria de política pública, com sua miríade de atores, uma intervenção pode estar associada a várias teorias de intervenção. Por razões pedagógicas, no entanto, este capítulo irá se afastar do pressuposto de apenas uma teoria de intervenção, inserida na própria intervenção pública institucionalizada, e envolvida com a forma como ela foi concebida para ser implementada e resultar em consequências em várias etapas. A teoria de intervenção é a própria espinha dorsal filosófica da intervenção. Neste sentido, não é uma teoria sobre os processos operacionais reais de implementação e resultados. Ela se refere a processos que *se destinam* a operar e que *são pensados* para dar um impulso à obtenção de resultados. Na sequência, o leitor deve manter essa limitação simplificadora acentuada em mente.

Caixa de Texto 3.1 Os termos “teoria de intervenção”, “teoria de programa” e relacionados

Uma das primeiras fontes de inspiração para o método de teoria de intervenção, o cientista social holandês Andries Hoogerwerf (1990), postula que a *teoria de políticas públicas*, termo de sua preferência, contém todos os pressupostos e suposições empíricos e normativos inseridos na intervenção inicial pelos seus autores. E as avaliações podem ser consideradas como testes de teorias de políticas públicas.

Teorias de intervenção vêm em diversas formas linguísticas. De acordo com Hoogerwerf, elas são referidas como “o modelo de impacto do programa” (NACHMIAS, 1979, p. 9 et seq.), “a teoria da ação do programa”, “o modelo do programa” (PATTON, 1990, p. 107) ou “o raciocínio subjacente ao programa”.

A “avaliação baseada em teoria” e a “avaliação orientada por teoria”, na maioria dos casos, referem-se à avaliação, utilizando a abordagem de teoria de intervenção. Mesmo “mapa cognitivo” parece

corresponder, pelo menos em parte, à teoria de intervenção. Ao cunhar o último conceito, Axelrod escreveu (citado de HOOGERWERF, 1990, p. 285 et seq.): “Um mapa cognitivo é uma forma específica de representar as afirmações de uma pessoa sobre algum domínio limitado, como um problema de políticas públicas (*policy problem*). Ele é projetado para capturar a estrutura de afirmações causais da pessoa e gerar as consequências que se seguem a partir dessa estrutura.” .

O termo comum é “teoria de programa”. No valor de face, no entanto, essa palavra denota teorias inseridas apenas nos *programas*, não em políticas, atividades, projetos, reformas, ou qualquer outra coisa que os governos fazem. Portanto, “teoria de intervenção” será o termo geral preferido aqui e “teoria de políticas públicas” e “teoria de programa” (BICKMAN, 1990 e CHEN, 1990) serão os termos usados quando as intervenções são pensadas como políticas, programas ou reformas.

Ocasionalmente, os avaliadores usam “teoria causal”. Como Hoogerwerf, afirma-se aqui que as teorias de intervenção não contêm apenas afirmações causais, mas também valorizam julgamentos. Além de caminhos causais (conexões) entre os componentes, uma teoria de intervenção sugere objetivos finais, que naturalmente são expressos em termos de frases valorativas (*value sentences*).

3.3 O Modelo Estendido de Sistemas como um Esquema de Análise da Avaliação

A abordagem da teoria de intervenção estimula avaliadores a interpretar, organizar e visualizar a *teoria de intervenção bruta* inerente à intervenção em um instrumento adequado para a avaliação continuada. Para este empreendimento, é necessário um sistema de categorias. Esse esquema, essa grade, é uma lente visual através da qual os avaliadores podem ver a filosofia embutida na intervenção. O esquema é mediação entre avaliadores e a teoria de intervenção bruta original. A grade conceitual é a adicionada pelos avaliadores pois, raramente, a teoria de intervenção bruta original é verbalizada e organizada de acordo com essa grade. O modelo de sistema é o melhor aparato conceitual que existe atualmente para a transformação de teorias de intervenção brutas em ferramentas estruturadas de análise.

O ponto de partida é o modelo sistêmico simples, consistindo de *componentes de input*, de conversão e de *output*, já apresentados dentro e em torno da figura 1.1 no capítulo 1. No capítulo 1, o modelo sistêmico simples também foi adaptado à avaliação de políticas públicas, dividindo a conversão em processo e estrutura, renomeando como administração, e adicionando resultado 1, resultado 2 e o resultado 3 a ele. Além disso, o leitor foi apresentado a resultados e implementação, bem como aos efeitos inseridos nos resultados. Repete-se a Figura 1.2 aqui como um ponto de partida para a criação de uma teoria de intervenção como uma ferramenta de avaliação.

Figura 3.1

Modelo Sistêmico com Administração, Output, Outcomes e seus Efeitos

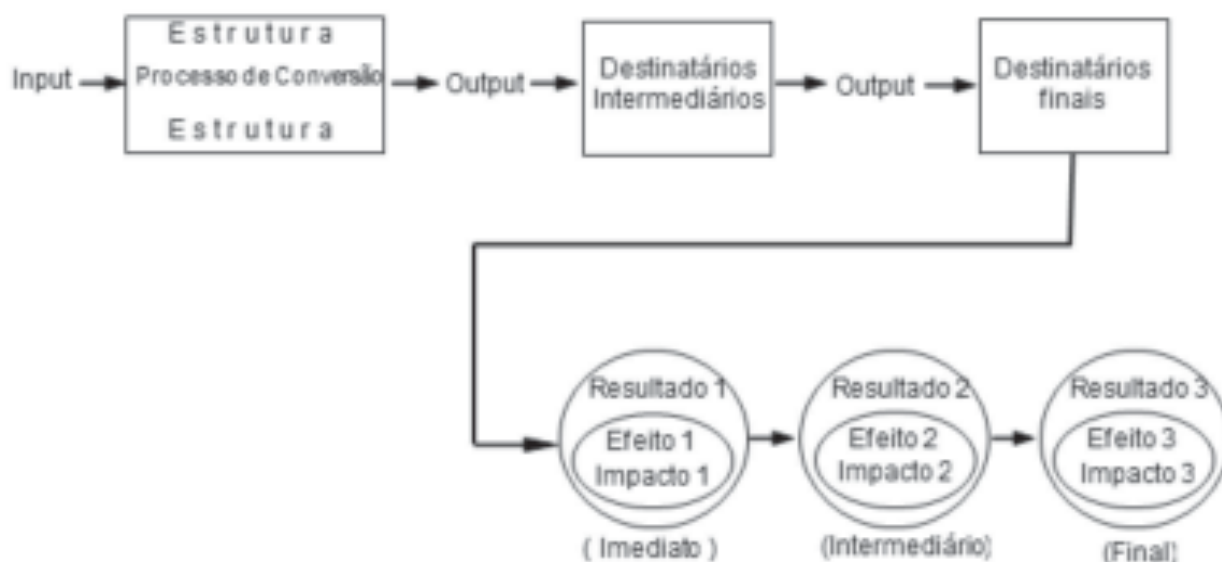


Além dos componentes mencionados, o modelo sistêmico na figura 3.1 apresenta também as *conexões entre os componentes*, desenhadas como setas causais. Agora, esse modelo pode funcionar como um esquema de categorias utilizável para a reconstrução de teorias de intervenção bruta. No entanto, mais algumas adições podem ser justificadas.

O modelo será mais fácil de usar, se alguns atores forem inseridos também. Primeiro, é razoável inserir o ator destinatário final após o *output*, mas antes do resultado 1. Destinatários finais são os agentes que a intervenção e a implementação devem atingir e influenciar no fim. Destinatários finais podem ser indivíduos ou coletividades, como famílias, creches para crianças, escolas e empresas. Palavras alternativas para destinatários finais são: grupos-alvo, alvos, clientes, participantes, usuários do serviço e assim por diante. Também pode

haver intermediários, ou mediadores, que a intervenção deve influenciar a fim de que estes, por sua vez, possam influenciar os destinatários finais. Em alguns programas voltados para escolas, por exemplo, instrumentos de governança são dirigidos aos pais (Destinatários intermediários), a fim de que eles influenciem seus filhos-alunos (Destinatários finais). Se esses dois tipos de agentes são inseridos, a extremidade distal do modelo sistêmico estendido aparece como na figura 3.2.

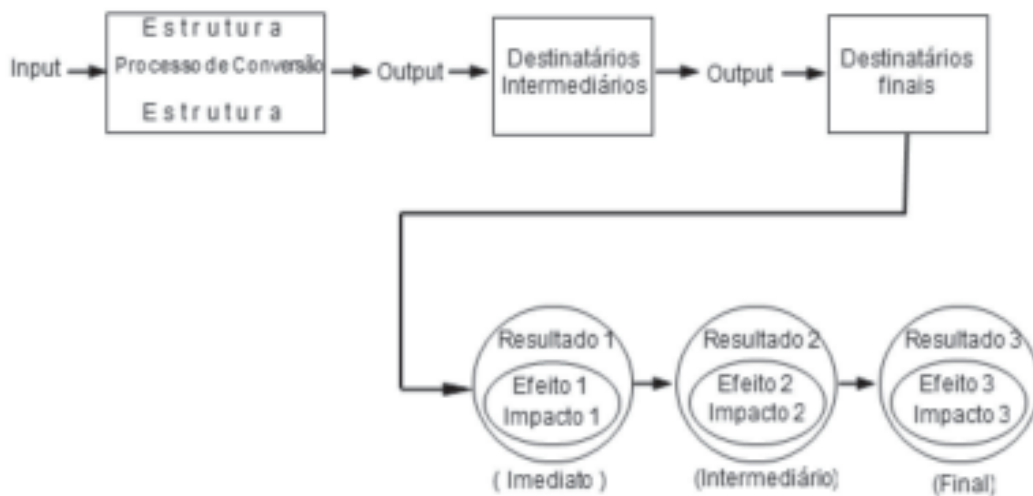
Figura 3.2
Modelo Sistêmico com Destinatários Intermediários e Finais



Processos administrativos (Administração) também, muitas vezes, contêm mais de uma camada de atores. Outros atores além dos órgãos públicos formais podem estar envolvidos, formando várias camadas de atores implementadores. No mundo de hoje, a administração pública é subcontratada a entidades privadas, tais como empresas, organizações de interesse público e organizações de interesse. Aqui basta exemplificar com duas camadas. Diferencia-se *gestores* (= o diretor em uma escola) de operadores de campo (= professores de uma escola)¹. Com esse aditamento, o modelo sistêmico toma a seguinte forma (Figura 3.3).

¹ No capítulo 1, vimos várias camadas de atores no caso da Ajuda aos refugiados afegãos (Cruz Vermelha em Genebra, Crescente Vermelho, em Islamabad, do Crescente Vermelho local nos campos de refugiados etc.)

Figura 3.3
Modelo de Sistemas com Gestores e Operadores



Aqui se deixa de abordar o modelo sistêmico na esperança de ter provado que ele pode ser expandido e se tornar mais e mais complicado, para converter-se em uma ferramenta mais útil de análise. Novamente, deve-se notar o que é o modelo de sistemas, e o que ele não é. Não é uma teoria de intervenção. Ele é uma ferramenta, um sistema de análise, para reconstruir hipóteses e suposições brutas, muitas vezes vagas e fragmentadas, inseridas na própria intervenção, em uma *teoria de intervenção refinada* que pode ser utilizada em avaliações futuras.

3.4 O Caso da Declaração de Combustível

Utilizando um programa real, será demonstrado que o método de teoria de intervenção esboçado pode ser aplicado em avaliação. Para realçar os pontos metodológicos, foi escolhida deliberadamente uma ilustração simples: as diretrizes da Agência Sueca do Consumidor (*Konsumentverkets*) para obter informações sobre o consumo de combustível de carros novos, a chamada Declaração de Combustível (*Bränsledeklarationen*)².

² No entanto, por razões pedagógicas, foi alterada a data da Declaração verdadeira de combustível para aproximadamente 40 anos à frente no tempo; em vez de no meio da década de 1970, presume-se que tenha sido adotada no ano passado e entrado em vigor em 1º de janeiro deste ano.

Suponha que um único avaliador foi contratado pela Agência do Consumidor para realizar uma avaliação sobre a implementação e os efeitos da Declaração de Combustível. Foi decidido que a abordagem da teoria de intervenção deve ser adotada. Será usada uma forma mais didática para explicar como a avaliação vai prosseguir. Primeiro, será interpretada e organizada a teoria de intervenção *bruta* incorporada na Declaração de Combustível, a fim de demonstrar como uma atividade tão interpretativa e organizadora deve ser realizada. Em segundo lugar, será mostrado como a teoria de intervenção *refinada*, assim reconstruída, pode ser usada em contextos contínuos de avaliação. O ponto de partida será com a teoria bruta narrada no texto da declaração. Para ir mais longe, será usado o trabalho de preparação e entrevistas com os participantes no processo de formação da Declaração. Também será implantado o chamado Princípio da Racionalidade.

No papel, a Declaração de Combustível afirma que todos os comerciantes de veículos novos de passageiros a partir de 2015 devem informar os compradores sobre o uso de combustível. Mais especificamente, os revendedores de devem fornecer informações sobre o consumo de combustível dos carros por cada 100 quilômetros de condução normal. A Declaração de Combustível deve ser uma regra suplementar (orientação, norma, *riktlinje*) para um estatuto regulamentar " a Lei de Marketing de 1976 (*marknadsföringslagen*), aprovada pelo parlamento nacional sueco. A última , a chamada lei-quadro (*ramlag*), contém uma breve exigência afirmando que na comercialização de diversos produtos ou serviços, as empresas podem ser obrigadas a fornecer informações importantes consumidores. Isso é tudo.

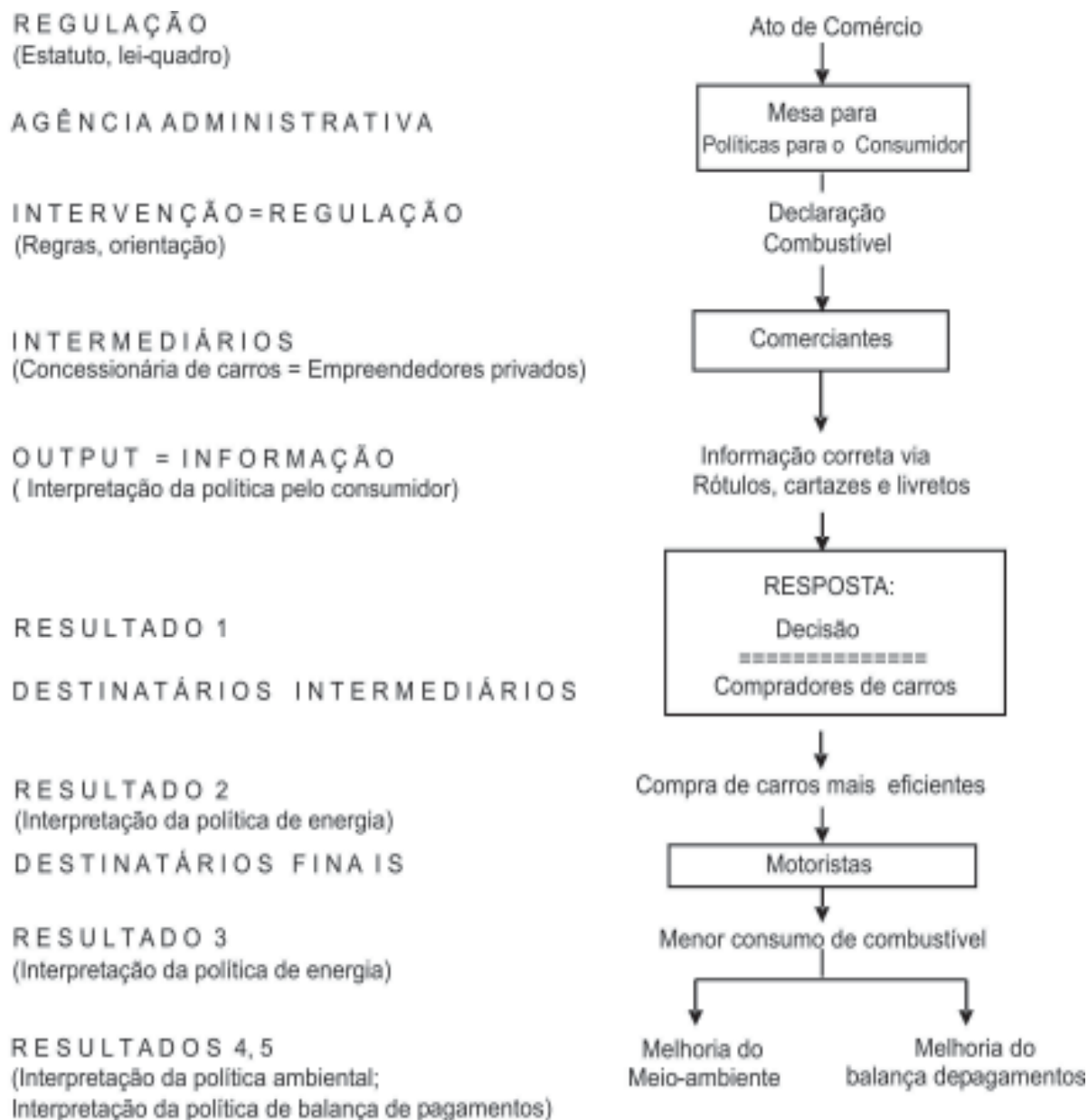
Na base dessa concisa recomendação contida na Lei de Marketing e nos trabalhos preparatórios que a precederam, a Agência Sueca do Consumidor – a autoridade do governo central responsável pela implementação do estatuto regulatório – promulgou muitas regras que especificam detalhadamente o tipo de informação que deve ser considerada de particular importância para o consumidor e, portanto, colocada à sua disposição. Existem regras para uma ampla variedade de bens e serviços, tais como fraldas, casas unifamiliares, viagens, vinho, cigarros, e serviços de restaurante. Todas as regras, incluindo a Declaração de Combustível, são casos de "abordagem de pacotes pormenorizados, com orientação pela mão" ("*the detailed-package*,

guiding-hand approach”), enquanto a Lei de Marketing é um exemplo da “alternativa da direção ampla, sem orientação” (“*the broad-direction, unseen-hand alternative*”), para usar a linguagem expressiva de Walter Williams (1976, p.276). A Declaração de Combustível é um pequeno *programa* detalhado entre várias centenas, desenvolvido no âmbito da *política* geral da Lei Marketing.

A informação deve estar em uma forma escrita e vem em três formas diferentes. Informações sobre o consumo de combustível devem ser oferecidas como uma *etiqueta no para-brisa* de cada carro individual. Um *cartaz* deve ser fixado na parede, exibindo as mesmas informações sobre todos os modelos de carros vendidos pela concessionária de automóveis. Finalmente, um *livreto* deve ser visivelmente colocado em algum lugar nas instalações, fornecendo as mesmas informações sobre todos os modelos de carros comercializados em todo o país. O rótulo, o cartaz e o folheto podem ser considerados *elementos básicos* do programa. Até agora, foi seguida a lógica, lendo e relendo a Declaração de Combustível na sua forma escrita.

O próximo passo no *desenvolvimento* de uma teoria de intervenção útil será refinar a teoria bruta, interpretando-a e reestruturando-a para o regime de categoria do modelo sistêmico (veja a Figura 3.1 e 3.2 acima), e, se possível, transformá-lo em um gráfico. O diagrama resultante aparece na figura 3.4.

Figura 3.4
**Um Primeiro Corte: a Teoria de Intervenção da
Declaração de Combustível Reconstruída**



O que a Agência do Consumidor deseja alcançar com a Declaração de Combustível? Evidentemente, a agência queria chegar aos compradores de carros novos (destinatários), agindo através de marketing, ou seja, através dos revendedores (administradores, implementadores). Mas que tipo de serviço a agência deseja fornecer aos compradores de veículos novos, e o que era esperado deles?

O objetivo pode ser o de oferecer ao cliente a possibilidade de ter conhecimento das informações corretas sobre o consumo de combustível de carros novos, ou seja, o *output* na figura 3.4. A ênfase está na *possibilidade*. O objetivo pode ser o de fornecer informações precisas e confiáveis através dos revendedores de automóveis, ponto. As forças de mercado cuidariam do resto.

Essa interpretação da política para os consumidores parece razoável. Afinal de contas, a Declaração de Combustível foi promulgada pela Agência do Consumidor, a autoridade pública encarregada de todos os programas nacionais de informação ao consumidor no país. E as intenções dos programas de informação aos consumidores suecos não são influenciar atitudes e ações dos consumidores. “Dê-lhes informações, mas deixe todo o resto para os consumidores”, esse é o lema. Nessa perspectiva, a Declaração de Combustível seria apenas um entre vários programas da política de consumidores geridos pela Agência do Consumidor.

No entanto, há algo mais com a Declaração de Combustível, algo que não é claramente definido em qualquer norma ou lei. Alguns elementos da cadeia de implementação pretendida não estão claramente definidos nos textos; talvez eles sejam tácitos e implícitos.

Ao complementar os textos de intervenção com os documentos preparatórios do processo de formação e com informações orais dos participantes do processo, descobre-se que a Declaração é adotada, não como um programa destinado ao consumidor, mas como uma medida de governança ambiental, de energia e de balança comercial. O objetivo é reduzir o consumo de combustível na condução de veículos (Resultado 3), que, por sua vez, é considerado como um meio para melhorar o meio ambiente através da redução das emissões de gases de efeito estufa e de substâncias tóxicas (Resultado 4), que, por sua vez, terá um impacto positivo sobre a balança comercial, graças à redução das importações de petróleo caro (Resultado 5). Consequentemente, o resultado final esperado é fundamental para melhorar o meio ambiente e favorecer a balança comercial sueca.

No entanto, como se pretende que esses impactos nos resultados finais sejam efetuados pela informação disponível nas concessionárias? Por essa razão, os textos são um pouco limitados e tácitos. Mais uma vez é necessário complementar os textos de intervenção com trabalhos preparatórios e outros documentos do processo de formação e com informações orais dos participantes do processo. Para preencher as lacunas também se deve recorrer ao raciocínio lógico.

Em primeiro lugar, as intenções devem ter sido já mais abrangentes com as informações nas concessionárias de veículos. A ideia deve ter sido a de que os possíveis compradores de carro devem ser induzidos a realmente decidir comprar carros que consomem menos gasolina do que teriam decidido sem essa informação (Resultado 1), e daí de fato comprar estes carros (Resultado 2).

E, em segundo lugar, um tipo adicional de destinatário deve ser inserido na cadeia, ao lado de compradores de automóveis, ou seja, os *condutores de automóveis*. O objetivo com a declaração é atingir os motoristas de carros através dos compradores de carros: os condutores são os *destinatários finais*, os compradores de automóveis são apenas *destinatários intermediários*. A compra de mais carros energeticamente eficientes (Resultado 2) era esperada para resultar no mesmo nível de condução de antes, de modo que o consumo de combustível entre o estoque de carros novos diminuiria (Resultado 3). O consumo de combustível reduzido, por sua vez, é considerado como um meio para atingir os dois resultados mais distais 4 e 5 anteriormente referidos.

3.5 Requisitos para Boas Reconstruções de Teorias de Intervenção Brutas

Uma boa reconstrução de uma teoria de intervenção deve reproduzir fielmente a teoria bruta. Deve ser uma representação fotográfica da mesma. A imagem na Figura 3.1 é uma *representação fotográfica da teoria bruta*? Evidentemente que não. As exigências de uma reprodução fiel nunca podem ser satisfeitas. Por quê?

Sem entrar nas incertezas envolvidas em toda interpretação (“traduttore, traditore”, como dizem os italianos), serão apresentadas duas razões. A primeira razão é assim: na estruturação da teoria bruta, o avaliador deve usar as categorias do modelo sistêmico como suas ferramentas. Uma vez que o avaliador codifica os dados brutos relevantes para o modelo sistêmico e sua estrutura **output — destinatários intermediários — outcome 1— outcome 2 — destinatários finais — outcome 3**, as informações serão interpretadas e organizadas em uma perspectiva favorável à avaliação de políticas públicas. *As categorias do modelo e da apresentação visual são adicionadas pelo avaliador*. Ele complementa um *esqueleto teórico*, uma *espinha conceitual*, um *esquema*, um *quadro*, que fornece ordem lógica para a teoria original. A isso dá-se o nome de *interpretação reconstrutiva*. Tal interpretação não é uma representação fiel do original.

Por que o avaliador deve selecionar exatamente o modelo sistêmico para destilar a partir de noções rudimentares uma teoria de intervenção refinada, ordenada? O objetivo é extrair uma lógica de ação, mostrando uma estrutura de pressuposições sobre atividades pretendidas, que, por intermédio dos atores, deve provavelmente causar certos efeitos. E, em avaliação de políticas públicas, o que interessa, muitas vezes, embora nem sempre, são as causas e consequências e as conexões de mediação entre elas.

A segunda razão contra a possibilidade de representação fiel é que partes importantes da teoria de intervenção tendem a ser tácitas. Esse é frequentemente o caso das conexões causais entre os componentes na teoria. Refere-se aqui às instâncias de decisão e aos chamados *mecanismos generativos*.

Na Declaração de Combustível, acredita-se que a emissão da Declaração afetará os comerciantes de modo a obrigá-los a exibir as informações prescritas, que, por sua vez, afetarão os compradores de automóveis que irão adquirir veículos mais eficientes no consumo de combustível, o que, por sua vez, irá induzi-los a não dirigir mais do que fariam se tivessem comprado carros que demandam mais energia. Mas o que quer dizer “afetar” e “induzir”? O que acontece exatamente quando uma intervenção “afeta” e, assim, traz efeitos? *Qual é a natureza da causalidade no mundo das intervenções governamentais?* (PETERSSON, 2006).

Considerando a questão a partir da perspectiva do ator afetado, uma intervenção governamental fornece *premissas* para as considerações e decisões. Premissas são aquelas condições às quais os tomadores de decisão geralmente atribuem importância nas suas decisões. Premissas são circunstâncias que os tomadores de decisão levam em consideração e escolhem para servir de orientação quando eles decidem. Esse raciocínio a respeito das premissas não é inteiramente satisfatório, mas o presente contexto impede que se entre em mais detalhes. (MATHIESEN, 2004).

Por isso, a ideia de premissas neste capítulo é levada a sério. Colocado de outra forma, refere-se ao *modelo de mecanismo generativo de causalidade*. O modelo generativo assume que existem mecanismos, que, em determinadas circunstâncias geram consequências. Estes mecanismos não podem ser diretamente observados. No entanto, eles existem.

Caixa de texto 3.2 Mecanismos generativos, premissas para a tomada de decisão.

A teoria de intervenção bruta da Declaração de Combustível afirma que informações precisas devem estar disponíveis nas concessionárias e que isso provavelmente levará a compra de mais carros eficientes em termos energéticos. São entre esses dois estágios que as *premissas* ou *mecanismos generativos* devem aparecer. Se a Declaração de Combustível tem efeito, o comprador de carro deve fazer considerações. Primeiro, ele deve ser exposto à informação, depois, prestar atenção nela, tornar-se interessado, entender do que se trata, adquirir novos conhecimentos ou ter antigos conhecimentos fortalecidos ou ativados por essa nova informação, mudar ou reforçar suas atitudes para com relação a carros eficientes e decidir comprar. Juntos, esses mecanismos, de 1 a 7, *constituem premissas ou mecanismos generativos*. Premissas ou mecanismos generativos consistem em intenções, objetivos e considerações dos atores. Os indivíduos que entram em um estabelecimento comercial de veículos têm que passar por essas fases, a fim de que a Declaração de Combustível tenha o efeito de aumentar a compra de carros mais eficientes em termos energéticos.

Mecanismos generativos também ocorrem em outras partes da cadeia. Os profissionais de marketing, ou seja, os revendedores de automóveis, também devem passar por essas etapas, para que a Declaração de Combustível garanta que as informações em rótulos, cartazes e folhetos sejam precisas e sejam exibidas nas concessionárias. O mesmo acontece com os motoristas de carro. Apesar de não ser explicado em detalhes na figura 3.1, os mecanismos generativos são indicados por duas pequenas setas para cima a partir dos atores pertinentes.

O movimento contemporâneo de busca de evidências, na sua forma mais radical, recomenda a experimentação randômica duplamente oculta como o projeto mais adequado para determinar a causalidade. Mas experimentos randomizados só podem provocar consequências, não mecanismos generativos (premissas) que podem conectar efeito com a causa na situação experimental.

Para trazer à luz mecanismos ou premissas ou partes de argumentos que são tácitos, mas que razoavelmente devem ter estado lá, o avaliador pode utilizar o Princípio da Racionalidade.³ O *Princípio da Racionalidade* implica que os agentes envolvidos supostamente agirão “adequadamente ou apropriadamente” na situação em que eles estiverem. (POPPER, 1985, p. 359). “O princípio de agir de forma adequada para a situação (...) é conhecido na literatura com o nome de ‘princípio de racionalidade’”, escreve Popper, e acrescenta: “um nome que tem levado a inúmeros mal-entendidos.” O ponto é que o Princípio da Racionalidade deve ser usado como um *postulado metodológico*, ou seja, um instrumento, uma ferramenta. Como uma ferramenta que será útil na criação de uma base, a fim de preencher, interpretar, as características tácitas ou vagas da teoria de intervenção bruta, não para explicar por que elas estão lá (HERTTING, 2003, p. 138 et seq.; BENGTTSSON, 1998).

“Agentes sempre agem de uma forma adequada à situação em que se encontram.” Isso, argumenta Popper, é “o princípio de racionalidade (...) na sua formulação mais fraca” (p. 360). Mas, mesmo nessa sua formulação mais fraca, a proposição contida nessa frase é obviamente falsa. Evidentemente, as pessoas não agem racionalmente nesse sentido, afirma Popper, e deve-se concordar com ele. Uma vez que se sabe que eles não agiram racionalmente, a proposição é falsa.

Mas então, qual é exatamente o status do Princípio da Racionalidade como uma ferramenta para a reconstrução das teorias de intervenção brutas ou para algum outro propósito nas ciências sociais? É uma aproximação razoável da verdade, diz Popper. Mas não é o seu valor verdadeiro que deve ser posto em causa no trabalho de reconstrução do avaliador. Pelo contrário, é a análise da situação que pode ser feita mais extensa, utilizando o Princípio como uma ferramenta. Ao usá-lo como ferramenta, é possível conhecer muito mais sobre as percepções e pressupostos da intervenção do que colocando sua verdade em questão (POPPER, 1985, p. 361 et seq.).

³ Interpretação reconstrutiva é recomendado por Chen (1990, p. 65 et seq.). O método já foi introduzido e avaliado em VEDUNG, Evert - **Political Reasoning**, (1982b, p. 114 et seq.),, em **Det rationella politiska samtalet** (1977, p. 85 et seq.) e em **Systematic Interpretation and Charity** (1983b).

A filosofia que está por trás dos esforços de reconstrução do avaliador pode ser resumida da seguinte forma. Se houver falhas ou partes potencialmente tácitas da teoria de intervenção provocadas por textos e outros dados disponíveis, o avaliador não para por aí e conclui que a teoria de intervenção é irracional nesses aspectos. Em vez disso, ele continua a procurar. Decidido a trabalhar a partir do pressuposto de que a teoria de intervenção bruta é racionalmente fundamentada, ele se pergunta: o que é que deve existir na teoria bruta para que ela seja sempre instrumental em relação à análise global da situação até agora montada? Embora tácito, quais os caminhos causa-efeito que deveriam lá estar? Graças ao Princípio da Racionalidade, o avaliador pode extrair e sintetizar relações consistentes (conexões) entre objetivos e meios que não são totalmente articulados, mas podem ser inferidos como razoavelmente perto da verdade.

Sabe-se que o pressuposto de racionalidade é falso, mesmo em sua formulação mais fraca, mas também sabe-se que ela tende a se aproximar da verdade. No entanto, avaliar a verdade da premissa de racionalidade não é o ponto. O objetivo é usá-lo como uma ferramenta. Dessa forma, presume-se que ela permite ao avaliador complementar as partes tácitas da teoria de intervenção bruta enunciada, por exemplo, no que se refere aos mecanismos generativos.

Para resumir. O avaliador tenta transformar os pressupostos e intenções brutas incorporadas na intervenção a respeito de como a intervenção deve trabalhar em uma teoria de intervenção *refinada*:

- 1) ao fazer uma leitura atenta dos documentos da intervenção;
- 2) através do recolhimento de informações, documentos preparatórios e entrevistas com os atores que participaram do processo de formação;
- 3) com a introdução do Princípio da Racionalidade (se necessário), a fim de compreender e apresentar os pressupostos tácitos e caminhos causais;
- 4) estruturando a interpretação que resulta em classes, tais como a implementação, resultados, metas, *outcome 1*, *outcome 2* e *outcome 3* extraídos do esquema de categoria do modelo sistêmico estendido; e
- 5) transferindo o resultado de tudo isso em exibição visual.

Caixa de texto 3.3. Doutrina de interpretação de Popper: construindo uma posição que vale a pena atacar

O uso de uma fraca presunção de racionalidade na reconstrução das teorias de intervenção, por exemplo, respeitante aos mecanismos generativos, assemelha-se ao *Principle of Charity* (princípio da caridade) na argumentação analítica. Em "*The Open Society and Its Enemies*" (A sociedade aberta e seus inimigos. Itatiaia Editora, 1998, 2 vols.) Karl R. Popper defende (1962, p. 246-293) o uso desse princípio. Uma interpretação, argumenta ele, deve ter a forma de uma reconstrução racional, e deve ser sistemática, deve tentar reconstruir o pensamento do filósofo como um edifício consistente... Devemos começar com a suposição de que um grande filósofo não é suscetível de estar sempre contradizendo a si mesmo, e, conseqüentemente, onde há duas interpretações, uma das quais farão [ele] consistente e outra inconsistente, esse filósofo vai preferir a primeira à segunda, se for razoavelmente possível.

Popper recomenda, assim, uma reconstrução racional do texto, para torná-lo consistente. A questão não é uma fixação ansiosa para os significados do cotidiano ou os significados técnicos dos documentos. O ponto de partida fundamental é que as filosofias de grandes mentes devem ser percebidas como livre de contradições. Essa suposição de consistência torna-se um argumento para eliminar as contradições entre os argumentos individuais. Na introdução de "*The Poverty of Historicism*" (A Miséria do Historicismo, EDUSP, 1980) Popper (1961, p. 3) proclama a famosa conclusão: "*Espero que, desta forma, tenha erigido uma que realmente valha a pena atacar*" (grifo dos autores).

3.6 Como a teoria de intervenção (refinada) pode ser usada?

Quando a teoria da intervenção bruta é interpretada, estruturada, descrita em forma narrativa, e visualmente retratada em um diagrama, é hora de refletir sobre como o resultado refinado pode ser usado pelo avaliador.

Todo esse trabalho de reconstrução pode ter ajudado os membros da equipe de avaliação a construir um entendimento compartilhado e muito mais profundo do que é a intervenção (DONALDSON, 2007).

Em segundo lugar, a equipe pode usar a teoria reconstruída para ampliar e aprofundar conhecimentos de várias partes interessadas para a situação de governança vigente. Nessa fase, é suficiente que a reconstrução seja comunicada de modo que ela possa ser digerida e discutida. A ideia seria criar uma maior *clareza* entre as partes interessadas, sem desencadeá-las para a ação. A teoria de intervenção cumpre um “*enlightment purpose*” (propósito de esclarecimento) (OWEN; ROGERS, 1998, p. 44). As partes interessadas mais afetadas pela Declaração de Combustível incluem a Agência do Consumidor, a sua gestão e seu pessoal, os revendedores de automóveis locais, os fabricantes de automóveis, os compradores de automóveis, e suas respectivas associações comerciais.

Em terceiro lugar, a teoria de intervenção reconstruída pode gerar *reflexões teóricas mais amplas* ao construir a formação, a implementação, a entrega do objeto de avaliação (“*evaluand*” em inglês), etc., como casos de algo mais geral. Ainda assim, o ponto é o de ampliar e aprofundar as perspectivas das partes interessadas e melhorar seus entendimentos conceituais a respeito das atividades.

Para começar, a Declaração de Combustível é um caso de *divulgação obrigatória*, uma estratégia de governança que se utiliza da regulação para trazer um fluxo de informações. As ferramentas finais – rótulos, cartazes e folhetos – são informativas, porque se espera que a informação crie o efeito sobre os prováveis alvos: os compradores de carros. A camada de nível médio da regulação, a Declaração de Combustível em si, deve obrigar os concessionários (regulação) a fornecer informações para os compradores de automóveis (informação). E a Declaração de Combustível, por sua vez, é emitida sob o mais alto nível do Marketing Act (regulação). Isso mostra que as autoridades públicas podem usar múltiplas camadas hierárquicas de instrumentos de governança, a fim de atingir as metas adequadas. Instrumentos de governança vêm em pacotes hierárquicos (VEDUNG, 1998a, p. 40 et seq.; BEMELMANS-VIDEC; RIST; VEDUNG, 1998b, p. 257 et seq.).

Além disso, os potenciais compradores devem encontrar essa informação através de *intermediários não-governamentais*: os revendedores de automóveis. Nesse caso, as concessionárias privadas de veículos suecas foram incumbidas com um dever público. Elas são obrigadas a atuar como operadores de nível de rua (*street-level operators*), divulgando informações ordenadas pelo Estado.

O papel dos revendedores no sistema é um exemplo de uma tendência na administração pública moderna de governo por *procuração*: dependência de agentes não-governamentais ou de terceiros para realizar fins públicos (SALAMON, 1981 e 2002; SALAMON; LUND, 1989).

Em quarto lugar, as intervenções podem ser classificadas como políticas públicas, programas, projetos e elementos (SHADISH Jr; COOK; LEVITON, 1991, p. 107). A partir dessa perspectiva, a pequena Declaração de Combustível de uma página acaba por ser um *programa* nacional entre muitos, deveria ser executada em vários locais (programa *multi-site*), ou seja, aproximadamente 1.000 concessionárias de veículos (projetos) desde a cidade mineradora de Kiruna, acima do Círculo Ártico, até Ystad, no extremo sul na costa do Báltico. Cada um desses projetos contém três *elementos*: rótulos de para-brisa, cartazes na parede e folhetos em algum lugar nas instalações. E, finalmente, a Declaração de Combustível em si, teve origem em uma *política* parlamentar, uma lei-quadro: o Marketing Act de 1976.

Quinto, a comunicação da teoria de intervenção refinada para os interessados mais afetados também pode desencadear *ações corretivas práticas*. Depois de deliberações e discussões, onde algumas pessoas argumentam contra e outros a favor da interpretação dada pelo avaliador a respeito da teoria, algumas partes interessadas podem tomar medidas práticas para fortalecer o desempenho real de algum componente no sistema de governança. Comerciantes de automóveis, por exemplo, podem começar a agir de forma mais vigorosa, se esforçando e aumentando a visibilidade da informação requerida. Contudo, isso também pode incitar à ação na direção oposta. Para algumas pessoas envolvidas, a teoria inserida no programa pode parecer tão irracional que a implementação deve ser desacelerada ou deixada cair no esquecimento. Em ambos os casos, a mera apresentação da teoria de intervenção para as partes interessadas pode iniciar uma fase de deliberação, o que pode ter repercussões práticas.

Sexto, a teoria de intervenção pode se tornar uma ferramenta prática no *planejamento, na execução, e na apresentação de trabalhos avaliativos de campo*. Ela pode fornecer ordem e estabilidade à coleta e à estruturação de dados e, assim, auxiliar na obtenção de uma imagem mais clara de como as coisas realmente se desenvolveram em certos aspectos importantes. Com base nessas ideias, monitores e avaliadores podem usar o modelo de teoria de intervenção de categorias para organizar o seu próprio relatório.

Finalmente, a teoria de intervenção é útil para a criação de um *sistema de monitoramento e avaliação* das atividades de governança. Estruturadores de sistemas de monitoramento devem trabalhar com *indicadores*. Uma teoria de intervenção pode mostrar exatamente onde, em uma cadeia de eventos, indicadores devem ser desenvolvidos. No caso da Declaração de Combustível, indicadores de consumo de combustível dos automóveis zero km comprados, do trabalho de transporte dos carros novos, das emissões de produtos químicos provenientes da combustão do combustível em carros novos, e do progresso da balança comercial nacional são necessários. Da mesma forma, são necessários indicadores sobre a exibição de rótulos, cartazes e folhetos nos estabelecimentos comerciais de automóveis. Posteriormente, os dados coletados sobre esses indicadores por meio de monitoramento podem ser usados para avaliar se a Declaração e sua implementação tiveram qualquer efeito sobre os resultados.

3.7 Teoria Substantiva, Teoria de Implementação

Teorias de intervenção consistem em componentes e conexões. Além dos componentes do modelo sistêmico, como produtos, resultados e as conexões causais (caminhos) entre eles, existem várias outras formas úteis de categorizar elementos das teorias de intervenção. Aqui só será tratada uma das formas, aquela originada de uma distinção entre a *falha na teoria* e a *falha na implementação*. Se a política não funciona, é devido ou à teoria substantiva defeituosa ou à execução deficiente. Seguindo essa orientação, uma teoria de intervenção pode ser dividida em duas partes:

- 1) teoria substantiva (de mudança);
- 2) teoria de implementação (de mudança).

Não há nenhuma fronteira natural entre essas duas partes de uma teoria de intervenção⁴. A linha deve ser desenhada pelo avaliador. E onde desenhá-la depende de como a *pergunta de avaliação* é formulada. Mais uma vez, deixe a Declaração de Combustível fornecer uma ilustração!

⁴ Alguns teóricos concordam que tanto a teoria substantiva quanto a teoria de implementação pertencem à teoria da intervenção. No entanto, alguns argumentam que a teoria do programa abrange apenas a teoria substantiva; a teoria de implementação (teoria da ação) é outra coisa que não pertence à teoria do programa. No início, Donaldson (2007, p. 13 e p. 23 et seq.), por exemplo, defende os dois lados antes de incluir ambos (teoria do processo de programa e teoria do impacto do programa) em teoria do programa.

Caso 1

A teoria substantiva, expressa em linguagem simples

Suponha que a questão de avaliação diz respeito ao impacto do tipo de instrumento de política envolvida (informações sobre a economia de combustível dos carros novos) no momento que consumidores comprem um carro novo. Aqui, a informação é causa, e a compra do carro pelo consumidor é o efeito. Em seguida, a teoria substantiva completa pode ser enquadrada de uma forma subsequente:

A teoria substantiva, expressa em formato se-então.

Se a informação sobre a economia de combustível dos carros novos é exibida nas instalações das concessionárias de veículos, *então* os compradores de automóveis vão comprar carros mais eficientes em consumo de combustível do que teriam feito de outra forma, todo o resto constante (*ceteris paribus*).

Teoria de implementação, linguagem simples

Nesse caso, a implementação será o processo desde a adoção da intervenção em nível nacional até a exibição dos rótulos de para-brisa, cartazes, folhetos em cada uma das várias centenas de concessionárias de veículos individuais.

Teoria de implementação, formato se-então.

1a) *Já que* o Estado tem adotado uma intervenção que requer que as concessionárias de automóveis exponham informações sobre o consumo de combustível dos novos automóveis, *então* as concessionárias provavelmente atenderão a essa prescrição, interessar-se-ão por ela, buscarão compreendê-la, submeter-se-ão a ela e decidir-se-ão por expor essas informações;

1b) *Se* as concessionárias de automóveis *decidiram* expor essa informação, *então* eles enviarão as informações pertinentes (rótulos de para-brisa, cartazes e folhetos) produzidos e distribuídos a eles;

2) *Se* os concessionários tiverem arranjado para que as informações pertinentes (rótulos de para-brisa, cartazes e folhetos) sejam produzidas e distribuídas a eles, *então* eles também *irão verdadeiramente exibir* a informação nas dependências de suas concessionárias.

3.8 Múltiplas teorias de intervenção

Nenhum teórico de avaliação afirma que há uma e apenas uma teoria de intervenção por intervenção. Pelo contrário; cada intervenção pode conter várias teorias. Em primeiro lugar, pode haver crenças conflitantes incorporadas na intervenção em si sobre como *se pretende que* a implementação *se desenrole* no longo do curto prazo, no médio prazo e no longo prazo dos resultados. Essas teorias de intervenção podem conter convicções a respeito de como a implementação está prevista para se desdobrar, quais *outputs* são projetados para acontecer, e quais são as consequências imediatas, intermediárias e finais desejadas e previstas para acontecer. Essas crenças podem ser adotadas pelo conselho da agência, pelo diretor-geral, pelos gerentes de nível médio, pelos políticos tomadores de decisão, pela oposição política, pelos membros do grupo-alvo, pelas organizações-alvo, e pelos distribuidores de equipamento técnico. Operadores individuais também podem considerar suas próprias teorias de intervenção.

Em segundo lugar, todas as partes interessadas mencionadas podem abrigar diversas teorias a respeito de como a intervenção *realmente funciona*. Em terceiro lugar, uma teoria de intervenção pode ser a *interpretação ponderada de um conjunto de partes interessadas* sobre a forma como uma atividade de governança provavelmente irá se desenvolver. A interpretação integrada pode ter sido criada pelos esforços de uma equipe de avaliação em induzir vários grupos a sentarem juntos para uma longa deliberação. Muitos, se não a maioria dos avaliadores, adotam essa perspectiva metodológica da teoria de intervenção. Em quarto lugar, *as pesquisas acadêmicas* podem apoiar ou levantar dúvidas sobre como uma determinada parte de um empreendimento do governo deveria trabalhar.

Todas essas teorias de intervenção podem constituir o ponto de partida da avaliação continuada.⁵

⁵ Scriven (2003) discute o papel tripartite de teorias de programa: teoria alegada de programa (*the alleged program theory*), a lógica real de programa (*the real logic of the program*) e teoria ótima de programa (*the optimal program theory*).

3.9 Limitações do método

O método de teoria de intervenção tem limitações. O risco de inobservância por parte do avaliador a respeito dos *efeitos colaterais imprevistos* nas fases de implementação, do produto e do resultado é a principal limitação. Os efeitos colaterais não previstos na teoria bruta incorporada não podem, naturalmente, ser reconstruídos por qualquer avaliador, por mais inteligente, em uma teoria de intervenção que aspira a representar essa teoria bruta particular. Se os efeitos colaterais e os caminhos causais explicativos não estão incluídos na versão reconstruída, o avaliador não vai ter nenhuma ajuda do método de teoria de intervenção em sua busca por efeitos colaterais. É aí que reside uma séria limitação potencial do método, já que os efeitos colaterais imprevistos sempre ocorrem, e também agora é uma tarefa de avaliação importante acompanhar esses efeitos colaterais.

A Declaração de Combustível como reconstruída neste capítulo *não contém explicações sobre os obstáculos e barreiras* nas suas fases de implementação e de resultados. Se uma avaliação do processo conclui que apenas 50% das concessionárias de veículos apresentam informações precisas, é natural querer deixar claro de que fatores isso depende. Da mesma forma, verifica-se que apenas 20% dos compradores que estão expostos à comunicação a seguem, o avaliador deve descobrir o porquê. Ou se a esperada redução do consumo de combustíveis não ocorrer, o avaliador deve tentar descobrir o porquê. Algumas teorias ou fatores que explicam tais desvios dos pressupostos da teoria de intervenção não estão ilustrados nas teorias de intervenção como interpretadas aqui (Figura 3.1). A esse respeito, elas devem ser complementadas por mais teorização, mais especificamente, sobre as causas. Em outras palavras, o avaliador não pode executar o seu trabalho com apenas uma teoria de intervenção.

A terceira limitação está relacionada com *fatores contextuais* que *facilitam* a implementação pretendida e evitam o surgimento de obstáculos. Tais fatores podem ser chamados “moderadores” (DAHLER-LARSEN, 2001). Moderadores são pré-requisitos no contexto em torno da situação de governança que devem ser atendidos para que a implementação, o *output* e o *outcome* funcionem como pretendido. Os moderadores podem ser incorporados na intervenção efetiva, e,

assim, incluídos na teoria de intervenção reconstruída. No entanto, não foi inserido nenhum na Figura 3.1. Se eles não surgem nem na intervenção, nem na percepção das partes interessadas de como a intervenção funciona, o avaliador deve tomar conhecimento dos mesmos, desde que esclareça algo em sua avaliação. Deve acrescentar-se como uma coisa natural que fatores contextuais também podem *impedir*.

MODELOS DE AVALIAÇÃO: UM RESUMO

Um modelo de avaliação é um *molde ou um padrão* que indica, em termos gerais, como uma avaliação pode ser conduzida. Tais moldes especificam como as avaliações podem ser realizadas em uma, duas ou no máximo três dimensões preferencialmente. A seguir, os modelos de avaliação, ou moldes, são organizados e discutidos com base em apenas duas dimensões: *a dimensão objeto* - que se refere a quais aspectos da intervenção avaliada os modelos se dirigem e *a dimensão valorativa* - que se refere a qual critério de mérito utilizado pelos modelos para avaliar o valor da intervenção sob avaliação. Para realizar a avaliação, as descrições dos objetos a serem avaliados são necessárias, mas não suficientes; os avaliadores também precisam de um critério para valorar o mérito dos resultados.

4.1 Modelos Substantivos e Modelos Econômicos

Suponha uma campanha municipal antitabagismo que tenha como alvo os estudantes do ensino médio de seis diferentes escolas, cuja meta seja que cinco por cento dos estudantes já fumantes devem deixar de fumar em três meses após a conclusão da campanha. Suponha ainda que duas avaliações são lançadas quatro meses após o final da campanha: uma para descobrir se a campanha atingiu seus objetivos (quantos por cento dos estudantes fumantes pararam de fumar dentro dos três meses após a conclusão da campanha) e a outra para avaliar o valor econômico da eventual realização de metas (realização de metas por unidade monetária gasta). Neste caso, a primeira avaliação usará o modelo de realização de metas; já a segunda utilizará uma variação dos modelos econômicos chamada de modelo de eficiência¹ (o custo da realização de metas). Possivelmente, os

¹ **Eficiência**: é quando algo é realizado da melhor maneira possível, ou seja, com menos desperdício ou em menor tempo. **Eficácia**: é quando um projeto/produto/pessoa atinge o objetivo ou a meta. **Efetividade**: é a capacidade de fazer uma coisa (**eficácia**) da melhor maneira possível (**eficiência**).

modelos de avaliação mais importantes do campo estarão incluídos aqui². Uma simples lista com as classes e subclasses é apresentada na Figura 4.1.

Figura 4.1
Modelos Substantivos e Modelos Econômicos exibidos
em lista simples e como esquema tipo árvore

I – Modelos Substantivos

A – Orientado à meta

- 1 Alcance de metas (eficácia)
- 2 Efeitos colaterais
- 3 Livre de metas

B – Orientado ao problema

- 4 Relevância

C – Orientado ao Ator

- 5 Cliente
- 6 Stakeholder
- 7 Colegiado
 - a. Crítica dos Pares
 - b. Autoavaliação

II – Modelos Econômicos

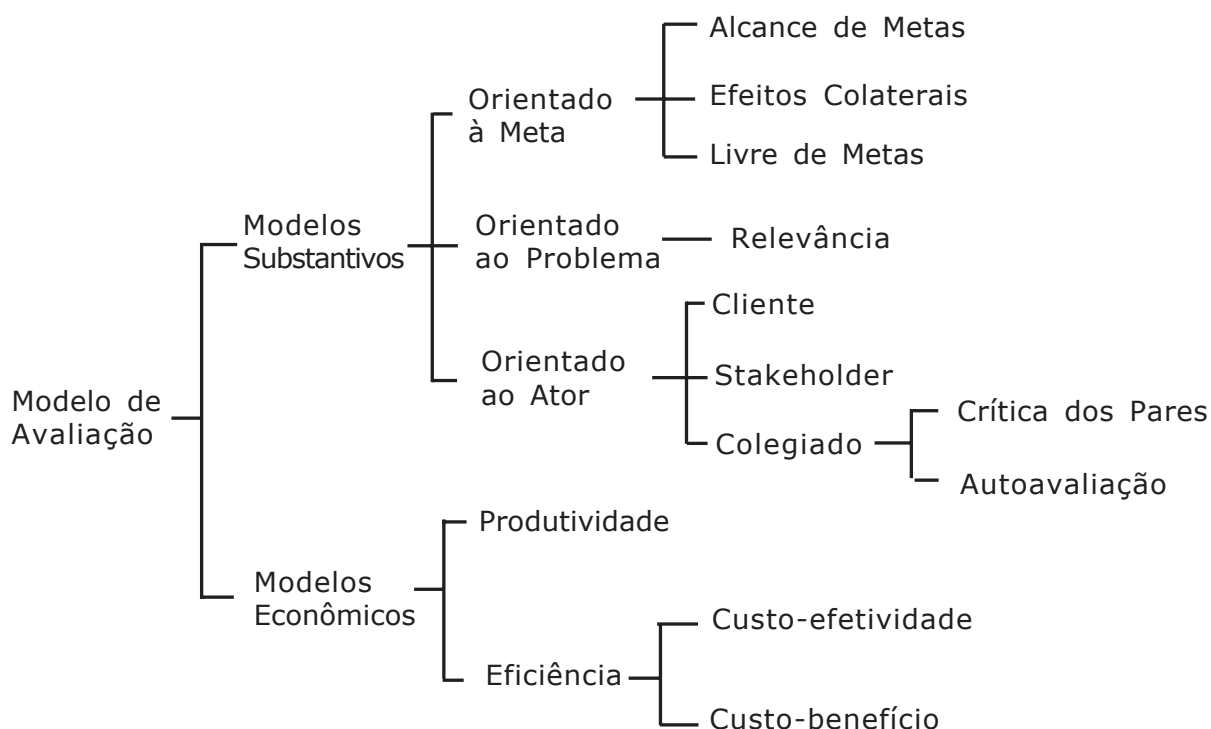
- 8 Produtividade
- 9 Eficiência: custo-efetividade; custo-benefício

Pedagogicamente, os modelos também estão exibidos em forma de árvore estendida na Figura 4.2

² Explicações de modelos de avaliação são fornecidos em Madaus, Scriven e Stufflebeam (1983), Stufflebeam, Madaus e Kellaghan (2000), Stufflebeam (2001, 22 abordagens) e (2004); Guba e Lincoln (1981), e House (1980). Em Shadish Jr., Cook e Leviton (1991), modelos de avaliação desenvolvidos por Weiss, Wholey, Scriven, Rossi, Cronbach, e Stake, são analisados. Stufflebeam e Shinkfield (2007, p.131-449) distingue 26 abordagens, onde cada abordagem é caracterizada em termos de dez (sic!) descritores (p.138); oito abordagens foram consideradas particularmente úteis (p.239 et seq.).

Figura 4.2

Modelos de avaliação como galhos, ramos e brotos de uma árvore



Nosso estudo se focará em nove modelos significativos, grosseiramente divididos em modelos substantivos e modelos econômicos. Enquanto os *modelos substantivos* (ou *apenas-substantivos*) prioritariamente abordam e julgam a intervenção como tal, seus resultados - *Outputs e Outcomes* - em termos de conteúdo substantivo (e, secundariamente, os processos que os precedem e os conectam), os *modelos econômicos* levam também em conta os custos. A última propriedade diferencia modelos econômicos dos modelos substantivos, sendo estes indiferentes aos custos.

Modelos Substantivos são subdivididos em abordagens orientadas à meta, orientadas ao problema e orientadas ao ator. *Modelos orientados às metas* são formados por uma abordagem bem conhecida e comumente utilizada (avaliação de alcance de metas, avaliação da eficácia), ou por uma abordagem bem conhecida, mas muito raramente utilizada (avaliação de efeitos colaterais), ou ainda por uma abordagem menos conhecida e raramente utilizada (avaliação livre de metas) no campo da avaliação. Este trio de modelos afasta-se dos objetivos da intervenção, isto é, objetivos incorporados na intervenção sob exame. Na avaliação alcance de metas isto é óbvio.

É igualmente óbvio em avaliação dos efeitos colaterais até onde efeitos principais e efeitos colaterais esperados, positivos ou negativos, estão envolvidos; quanto aos efeitos colaterais não esperados, é apenas indiretamente óbvio. Enquanto a avaliação livre de metas não usa metas como critério de mérito, o nome “Livre de Metas” indica que isto é, no mínimo, contrastado com metas em que possa se utilizar de não-metas *ex post* como critério de mérito.

Modelos de Relevância pertencem a um grupo especial de modelos não-econômicos orientados ao conteúdo chamados de *orientado aos problemas*. Eles não usam ou indicam metas como critério de mérito, mas, ao invés disso, apontam soluções aos problemas subjacentes.

Derivando seus critérios de mérito de vários atores (reais ou idealizados) envolvidos na formação e na implementação da intervenção, e não a partir da própria intervenção substantiva, as avaliações orientadas ao cliente, do *stakeholder* e por colegiados são modelos *orientados aos atores*.

Repetindo, a divisão primária, nas figuras 4.1 e 4.2, em modelos substantivos e econômicos é baseada na dimensão do objeto (apenas-substantiva versus apenas-substantiva + custo), enquanto as divisões secundárias se originam dos critérios de méritos utilizados em avaliações. No entanto, há várias outras possibilidades. Pode-se organizar a pesquisa (o levantamento) de acordo com a posição de *filosofia da ciência* em que ela se encontra, por exemplo, abordagens positivistas, construtivistas ou realistas, ou de acordo com a *metodologia de averiguação dos efeitos da intervenção*. Este último ponto de vista induzido pelos métodos é o preferido pelo atual movimento por políticas públicas e administração pública baseadas em evidências. Modelos são ordenados de acordo com a robustez dos delineamentos para aferição de impactos (experimentais, quasi-experimentais, séries temporais, rastreamento de processos, etc.). Apesar deste livro não ter organizado sistematicamente os modelos, os leitores encontrarão estes delineamentos em vários capítulos relacionados.

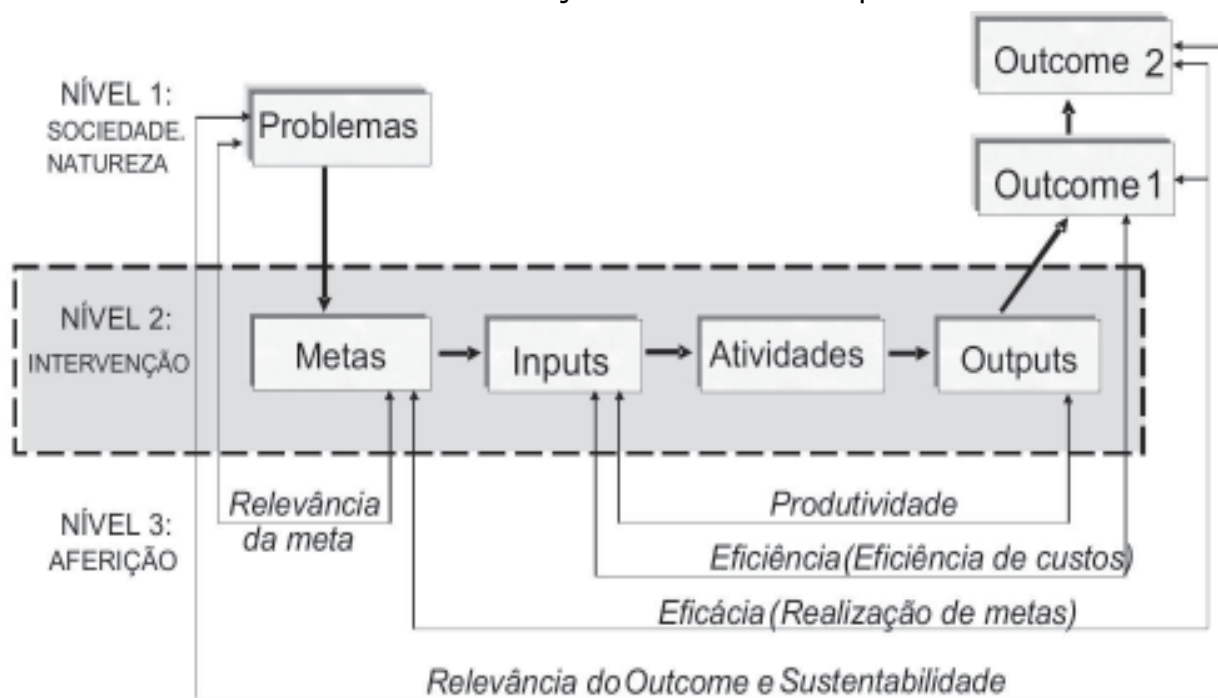
Há muitas possíveis dimensões que podem ser utilizadas para classificar modelos. *Métodos de Coleta de Dados* (métodos documentário, interrogativo, observativo, audio-visual) é uma; outra é *a questão não-econômica a ser avaliada* (processos e resultados;

políticas, programas, projetos, elementos e ferramentas de governança; governança colaborativa, administração orientada aos resultados, administração orientada ao processo e administração orientada ao cliente). Enquanto discutido ocasionalmente sob cada modelo de avaliação, nenhuma destas possíveis dimensões é utilizada como *fundamentum divisionis* para a classificação aqui adotada.

4.2 Modelo de avaliação da União Europeia

Vários modelos podem ser combinados na mesma avaliação. Um exemplo ilustrativo pode ser extraído a partir da União Europeia. Um dos moldes favoritos da Comissão Europeia recomenda o uso de quatro dos modelos apresentados aqui ou anteriormente (VEDUNG, 1997, p. 36) e também um quinto (sustentabilidade). Uma versão ligeiramente modificada deste rico modelo combinado é mostrada na figura 4.3.

Figura 4.3
Modelo de Avaliação da União Europeia ³



³ Para cobrir políticas ambientais da UE foi introduzida dimensão NATUREZA, além da Sociedade.

A força pedagógica do molde europeu vem de seus três níveis: nível 1) com Problemas e *Outcomes* envolvendo a Sociedade e a Natureza; nível 2) compreende a intervenção sob escrutínio; e nível 3) a Aferição, ou seja, a avaliação da intervenção. Também devemos destacar que o modelo sistêmico utilizado no molde (retângulos amarelos) permite estruturar o processo de intervenção em metas, conversão (*inputs*, atividades), *outputs* e *outcomes*. Além de produtividade e eficiência, os avaliadores da União Europeia devem focar na realização de metas (eficácia) e relevância (metas, bem como resultados, *outcomes*, obtidos em relação ao problema subjacente). Além disso, eles devem avaliar a sustentabilidade no sentido de longevidade dos resultados, depois que o suporte financeiro da União Europeia for cortado.

MODELOS DE AVALIAÇÃO SUBSTANTIVA

Este capítulo levará o leitor em uma jornada com paradas para examinar nove modelos de avaliação. Primeiro visitaremos a abordagem orientada a metas e a orientada a problemas: os modelos de alcance de metas, efeitos colaterais, livre de metas e de relevância. Depois nos deteremos a outro conjunto de modelos unicamente de substância: as abordagens do cliente, do *stakeholder* e do colegiado compõem o grupo dos modelos orientados aos atores. No capítulo seguinte, a jornada terminará com o exame de dois modelos de economia: produtividade e eficiência.

5.1 Modelo de alcance de metas

Os dois ingredientes básicos da avaliação por alcance de metas são a mensuração da realização de metas e a aferição do impacto da intervenção¹. A pergunta chave na *mensuração da realização de metas* é: (1) Os resultados estão de acordo com as metas da intervenção? Isso, por sua vez, abarca duas atividades distintas: (1.1) a clarificação das metas da intervenção (a função das metas) e (1.2) a mensuração de fato da realização das metas premeditadas da intervenção (a função do cumprimento das metas) (LANE, 1987). O problema da *aferição do impacto* pode ser formulado: (2) Os resultados são produzidos pela intervenção?

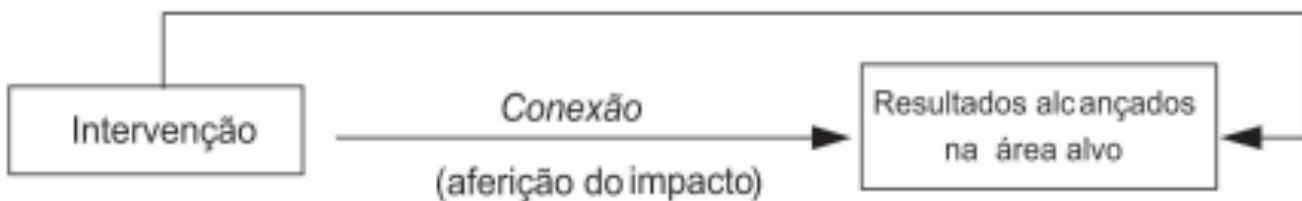
¹ Enquanto "avaliação do alcance de metas" (KAUFMAN; THOMAS, 1980, p. 126 et seq.) é uma expressão comum na literatura, o termo "avaliação do atingimento/obtenção de metas" (SCRIVEN, 1991, p. 178) é frequentemente utilizado também. Outras denominações incluem: "o modelo racional", "abordagem orientada aos objetivos" (GUBA; LINCOLN, 1981, p. x); ver também o modelo Tyler de avaliação em Guba; Lincoln (1981, p. 3 et seq.) e em House (1980, p. 27), "abordagem baseada nas metas" (HOUSE, 1980, p. 26 et seq.) e "a abordagem comportamental dos objetivos" (HOUSE, 1980, p. 26 et seq.).

O modelo de alcance de metas (i.e. *avaliação de efetividade, abordagem da efetividade*) é um ideal de simplicidade e lucidez. Seu primeiro passo consiste em identificar as metas da intervenção, destacando seu significado de fato e sua ordem hierárquica, transformando-as em objetivos mensuráveis, e determinando até que ponto são compreendidas. O segundo passo implica em averiguar o grau em que a intervenção promoveu ou dificultou o atingimento das metas.

A simples anatomia da avaliação por alcance de metas é exposta na figura 5.1:

Figura 5.1
Avaliação de alcance de metas (avaliação de eficácia)

Os resultados alcançados estão de acordo com as metas?
(medição da realização de metas, monitoramento de resultados)



A avaliação por alcance de metas utiliza as metas premeditadas, iniciais, anunciadas do programa como critérios de mérito e de organização para a avaliação. Uma vez que programas são inaugurados para atingir certos alvos, parece natural aferi-los à luz desses alvos. Se os alvos são atingidos, então os programas terão sido bem sucedidos; se não, eles terão falhado. Ernest House (1980, p. 26) escreve:

[ela] pega as metas do programa conforme anunciadas e então coleta evidências de se ele atingiu essas metas. Essas metas servem como fonte exclusiva de padrões e critérios. O avaliador afere o que os desenvolvedores do programa dizem pretender atingir. A discrepância entre as metas anunciadas e os resultados é a medida do sucesso do programa.

Isso não significa, em termos práticos, que os avaliadores devam, antes de tudo, buscar pelas metas. O ponto crucial é que as metas são o ponto de partida *lógico*. Afinal de contas, a maior tarefa da avaliação, afirmam os defensores do modelo de alcance de metas, é determinar se as metas preestabelecidas do programa foram de fato atingidas e então tentar descobrir até que ponto o programa contribuiu para a realização da meta.

5.2 Prós e contras do alcance de metas

Surgindo nos anos 1950 e 1960 como uma inovação em políticas públicas, o modelo de alcance de metas reinou supremo. Nenhum outro modelo existia. A avaliação deveria ser científica e livre de valores, trazendo pesquisa social imparcial e não tendenciosa sobre os meios para se chegar a certos fins no que concerne a criação de programas e políticas públicas. Toda avaliação requer critérios de mérito pelos quais a intervenção possa ser avaliada. O modelo de alcance de metas, se argumentava, oferece uma solução objetiva e não enviesada ao problema do critério de mérito na avaliação. Uma vez que as metas da intervenção são explicitamente enunciadas nos documentos de intervenção originais ou no trabalho preparatório, elas podem ser estabelecidas empiricamente, através de um ato de interpretação livre de valores. Julgando os resultados da intervenção com base em metas estabelecidas por terceiros, o avaliador cuidadoso consegue evitar assumir uma postura pessoal e subjetiva a respeito dos méritos e deméritos da intervenção a ser avaliada. Ao passo que se as finalidades (metas) forem colocações de valor e, portanto, externas ao domínio da ciência, os meios para se chegar a essas finalidades (metas) serão colocações factuais que podem ser investigadas e averiguadas por métodos científicos. Dado que finalidades (metas) foram definidas por membros fora da comunidade científica e expressamente reconhecidos como subjetivos, acadêmicos e outros avaliadores independentes poderiam examinar vários meios para se chegar a esses fins externamente definidos de maneira empírica, objetiva e livre de valores. Essa visão do modelo de realização de metas se baseava na *racionalidade de meios e finalidades*, emanando do pensamento de Max Weber e do positivismo lógico. O argumento poderia ser chamado de *razão da objetividade científica* para o modelo de realização de metas (VEDUNG, 2010).

Desde o começo dos anos 1970, o modelo foi pesadamente atacado por esse e outros motivos ficou em segundo plano. Ainda assim, ele experimentou duas renascenças, a mais recente no tempo presente sob a bandeira das políticas baseadas em evidência ou informada em evidência. A primeira renascença, nos anos 1980, 1990 e 2000, veio acoplada à popularidade da gestão de resultados, às vezes sob o nome *avaliação de efetividade*. A justificativa disso foi a ascendência do *argumento da democracia representativa* fundado na *primazia da corrente de governança parlamentar*.

A palavra democracia deriva do grego *demos*, povo, e *kratein*, governar. Numa democracia, o povo governa – ponto. Ainda assim, devido à falta de tempo, o povo não pode participar em mais do que uma pequeníssima parte do volume enorme de decisões que precisam ser tomadas no que concerne ao bem-estar dos cidadãos. E a falta de competência torna impossível para as pessoas, em uma ponta do país, tomarem decisões sábias que competem à outra ponta acerca, por exemplo, da colocação de pacientes na fila de espera para uma cirurgia ou do cuidado dia após dia com um cidadão de mais idade doente em um asilo para os idosos, do setor público. Dessa forma, os cidadãos devem eleger representantes políticos, para que tomem as decisões em seu nome. Mas os representantes em assembleias políticas não têm tempo ou competência para tomar todas as decisões. Eles devem delegar seu poder a governos para que tomem suas decisões por eles. Mas os governos também não têm tempo ou o conhecimento específico necessário, de forma que eles, por sua vez, têm que delegar a tomada de decisões aos profissionais e aos servidores públicos. O setor público é composto de uma longa cadeia de relações principal-agente.

No sistema parlamentarista sueco, se uma agência adota um programa para alcançar certas metas professadas, essas metas são as metas da sociedade, porque o poder de tomada de decisão da agência foi delegado a ela pelo governo, e o governo recebeu sua autoridade para tal do parlamento, e o parlamento, por sua vez, dos cidadãos. É uma virtude decisiva que o modelo de alcance de metas reconheça esse aspecto democrático das metas do setor público.

Em suma, o modelo de alcance de metas marca um ponto decisivo a respeito da sua inclinação para a corrente de governança parlamentar e a democracia representativa, particularmente para a perspectiva leiga dos cidadãos. A perspectiva do cidadão foi

repetidamente reforçada nesse contexto porque ela é raramente aludida no discurso internacional sobre a avaliação².

Por outro lado, o modelo de alcance de metas sofre de fraquezas e falhas persistentes, para as quais se volta a atenção agora. Algumas razões gerais significativas contra o modelo de alcance de metas são³:

- o argumento da nebulosidade
- o argumento das agendas ocultas
- o argumento de cima para baixo
- o argumento dos efeitos não intencionais

O *argumento da nebulosidade* afirma que as metas da intervenção são critérios de mérito deficientes devido à sua obscuridade. Existem dois tipos de *obscuridade de metas*: indeterminação de meta e catálogos de meta. Para começo de conversa, programas são baseados em *metas indeterminadas*, “inexatidão terminológica”, para usar a divertida expressão de Winston Churchill. Metas particulares podem ser *ambíguas* e carregar dois ou mais significados simultâneos. Ainda assim, a ambiguidade nesse sentido de duplos significados é excepcional da linguagem burocrática e de políticas públicas, e pouco incomoda os avaliadores. Mais incerteza é causada pela *vagueza*. A meta é vaga se ela não delinea claramente os casos em que ela é ou não é aplicável. A fronteira externa que delimita a extensão de uma palavra vaga é tão confusa que, em um determinado intervalo, é impossível saber o que está incluído na extensão e o que não está. Rampante na retórica política, a vagueza é um expediente favorito para aplacar conflitos políticos com fórmulas semânticas sem realmente resolvê-los⁴.

² Ao elaborar sobre a democracia deliberativa, nem House (2004, p. 220 et seq.) nem Lincoln e Guba (2004, p. 239) mencionam a democracia representativa. Lincoln e Guba são claros: aqueles que lançam mão de design experimental ou *quasi*-experimental não são citados como buscando de alguma forma servir à democracia ou à democracia representativa, enquanto que aqueles que praticam avaliação construtivista com métodos dialógicos e amplo envolvimento dos *stakeholders* são “informados pela teoria social democrática”.

³ Para crítica do modelo de alcance de metas, ver Deutscher (1976) e Meyers (1981, p.110 et seq.). A exposição em Chen (1990, p.168 et seq.) é muito clara e interessante.

⁴ Para funções da nebulosidade muito bem apropriadas, consultar o capítulo sobre rastreamento de processos (*process tracing*) e estudos de caso.

A segunda grande obscuridade é gerada pelos *catálogos de metas*. A maior parte das grandes reformas sociais contém uma lista impressionante de metas diversas. Enquanto uma única meta pode ser tratada como a principal, geralmente também se defende que ela seja balanceada frente às outras, incluindo as potencialmente conflitantes. Mas os *trade-offs* necessários entre as várias metas não são indicados, o que torna impossível eleger de tamanha lista de metas um ou poucos resultados esperados distintos e transparentes. Assim, as metas do programa não oferecem qualquer guia seguro para a amálgama contínua de dados. Elas tampouco são lúcidas o bastante para serem usadas como critérios de mérito contra os quais medir sucessos, insuficiências e falhas.

O regime regulatório do governo sueco sobre a gestão florestal contém um catálogo de objetivos individualmente vagos e coletivamente contraditórios. O primeiro parágrafo da Lei de Gestão Florestal afirma que as florestas devem ser geridas de forma a: 1) assegurar a capacidade de produzir madeira. Entretanto, a lei também declara que: 2) proteção da natureza e 3) outros interesses da sociedade devem ser levados em conta. O negócio florestal deve operar com a devida preocupação pelas funções da floresta como 4) habitats para plantas e 5) animais, seu papel no 6) equilíbrio do sistema de águas e 7) clima local, para 8) vida ao ar livre e 9) recreação. Deve-se prestar atenção aos valiosos 10) ambientes culturais e 11) paisagens.⁵

Não se estipula exatamente quão alto deverá ser a produção sustentável de madeira para ser balanceada com os outros dez objetivos em várias situações específicas. Isso certamente é sábio, visto que seria difícil especificar nas decisões centrais como a pesagem deve ser feita em cada caso. Entretanto, essa sabedoria cria dificuldades para o modelo de alcance de metas, visto que o programa não estipula um ou vários resultados esperados.

Se o avaliador se apegar à noção de objetividade científica, princípio fundamental na avaliação por alcance de metas, ele não será capaz de trazer à tona, a partir das formulações gerais de objetivos, um ou vários resultados indiscutivelmente claros. Para chegar a tais objetivos, o avaliador provavelmente deverá esclarecer os objetivos do programa e estipular prioridades entre eles, o que implica sair da esfera de investigação social imparcial e entrar em uma área de especulação subjetiva.

⁵ Svensk Författningssamling (Swedish Codes of Statutes) (1979, p. 429), e Skogsstyrelsens Författningssamling (The Forest Act) (1986, p. 6)

O argumento da *nebulosidade* das metas revela um importante desajuste entre os requerimentos do modelo de alcance de metas e a forma como intervenções públicas e atividades são geralmente formuladas. Se as autoridades eleitas e outros envolvidos na aplicação da intervenção não especificaram metas individuais em objetivos mensuráveis, e se eles não balancearam os vários objetivos declarados em um ou vários razoavelmente transparentes resultados (*outcomes*) globais ou medidas de *output*, o avaliador utilizando o modelo de alcance de metas não poderá resumir suas descobertas em um julgamento completamente livre de valores. Ele poderá fazer isso apenas depois que tenha interpretado e esclarecido as metas e prioridades entre eles de forma que não lance dúvidas na objetividade de todo o empreendimento.

Um segundo contra-argumento sugere que o modelo de alcance de metas desconsidera o papel das *agendas ocultas* na formulação de políticas públicas.

Metas oficialmente expressas assemelham-se à ponta de um iceberg, como é sugerido. Constituem apenas o que os tomadores de decisão querem manter para o público. As metas estabelecidas, muitas vezes, têm um caráter simbólico, não destinadas a serem completamente atingidas, enquanto os reais motivos operacionais são tácitos e escondidos de vista, como a enorme massa do iceberg abaixo da água. Esses motivos reais apontam para outras direções.

A agenda política escondida por trás de uma privatização de companhias estatais pode não ser o aumento da eficiência para os consumidores, mas o fortalecimento da rede de poder que dá suporte para os partidos de centro-direita. *Agendas ocultas* também incluem motivos estratégicos. O ponto pode ser fortalecer a coesão partidária, atrair eleitores na próxima eleição, preparar o terreno para um governo de coalizão. Uma vez que essas metas estratégicas são atingidas, os políticos perdem o interesse em implementar as provisões substantivas da intervenção.

Para cientistas políticos, essa é uma análise básica das forças que motivam as reformas públicas. É claro que considerações estratégicas desempenham um papel importante nas políticas públicas. Mas qual é a importância dessa estrutura de agendas ocultas em relação ao modelo de alcance de metas? Os resultados também deveriam ser avaliados contra tais parâmetros estratégicos como

vencer as próximas eleições, fortalecer a coesão interna do partido, ou manutenção das coalizões governamentais? Nesse caso, temos que admitir, o modelo de alcance de metas, como interpretado aqui, seria insuficiente, visto que utiliza apenas as metas substantivas como parâmetros.

Por outro lado, avaliadores de alcance de metas podem legitimamente argumentar que as *agendas ocultas* podem muito bem ser consideradas nesse modelo. Elas podem ser utilizadas como fatores que *explicam* porque resultados substantivos ocorreram ou não ocorreram. Se, por exemplo, as metas substantivas do programa não foram cumpridas, não há nada na lógica do modelo de alcance de metas impedindo os avaliadores de oferecerem *agendas ocultas estratégicas* como uma das razões de o porquê disso ter ocorrido. Fornecer tais explicações deve ser considerada a maior atividade em pesquisa de avaliação, e é perfeitamente consistente com o modelo de alcance de metas.

Concluindo, o modelo padrão de alcance de metas pode considerar agendas ocultas implícitas nas decisões da intervenção, mas como um fator que explica os graus de realização de metas.

A *terceira* objeção, o *argumento de cima para baixo*, sustenta que o modelo de alcance de metas ocorre de uma *visão convencional irrealista da relação entre política e administração* de modo que a administração execute fielmente as decisões dos políticos. A avaliação de alcance de metas é acusada de tratar a implementação como mecânica, linear, e inteiramente controlada pelo topo.

Mas essa depreciação não é o ponto. Não há nada na lógica do modelo de alcance de metas obrigando-nos a acreditar que a intervenção produziu os resultados previstos. Pelo contrário, o eventual impacto da intervenção na área alvo é considerado um dos dois problemas a serem investigados em uma avaliação de alcance de metas. Consiste em um conjunto de *questões* a serem levantadas em uma avaliação, e não um conjunto de *respostas* a essas questões. A seta de “intervenção” para “resultado alcançado” na figura 5.1 mostra não como a *conexão* realmente é, mas sim o que o avaliador tenta descobrir em uma aferição de impacto do alcance de metas. Ele pode muito bem chegar à conclusão de que não há relação causal entre o programa e o resultado. A objeção de que a abordagem de alcance de metas trata a implementação como unilinear e de cima para baixo não tem influência alguma sobre a lógica do modelo.

A *quarta* objeção, sobre os *efeitos não intencionais*, é a mais importante. As intervenções no setor público invariavelmente levam à consequências que não foram previstas na situação da decisão original. “É difícil prever, principalmente quanto ao futuro”, como disse Niels Böhr em tom de brincadeira⁶. Se os avaliadores se limitassem a averiguar o cumprimento de objetivos preestabelecidos, a busca por efeitos colaterais inesperados não estaria incluída no processo da avaliação. As descobertas da avaliação exibiriam uma visão limitada dos eventos. Em toda probabilidade, uma intervenção que gere alguns *spin-offs* interessantes deve ser melhor do que uma que produza *spillovers* indesejáveis (FOSS HANSEN, 1989, p. 204).

Além disso, a prevalência de ações intencionais resultando em consequências colaterais não esperadas é uma forte razão – além dos efeitos negativos – para se fazer avaliações em primeiro lugar (MEYERS, 1981, p. 18 et seq.). Projetá-las de forma a não permitir a possibilidade de descobrir tais efeitos é pernicioso.

Suponha que alguém é encarregado de conduzir uma avaliação de alcance de metas do programa do Rei Fernando e Rainha Isabel de enviar Cristóvão Colombo para descobrir uma rota marítima pelo Oeste para a Índia. O avaliador deve concluir que o programa é um fracasso porque a Índia não foi alcançada e o objetivo do programa não foi cumprido. Não há razão para considerar o efeito colateral imprevisto que mudou a história do mundo, a descoberta da América, a criação dos impérios Espanhol, Português, Francês, Holandês e Britânico que se seguiu, e mais o enganoso e cruel extermínio de civilizações aborígenes, o aumento da escravidão dos negros, e assim por diante, visto que o modelo de alcance de metas não presta atenção aos fenômenos de fora da área alvo.

Em suma, por ser fundamentado na democracia representativa por tomar seriamente os objetivos oficiais da questão é a maior força do modelo de alcance de metas. Dessa perspectiva, os principais, como políticos eleitos e líderes de agências e instituições, precisam de avaliação de metas para verificar se seus agentes realmente fazem o que devem fazer. No entanto, o modelo tem problemas com objetivos nebulosos e catálogos de objetivos. A refutação mais convincente, entretanto, emana de sua cegueira aos efeitos colaterais imprevistos.

⁶ Niels Böhr (1885-1962), físico dinamarquês. Suas pesquisas sobre a estrutura atômica e física quântica fizeram com que recebesse o Prêmio Nobel de Física em 1922.

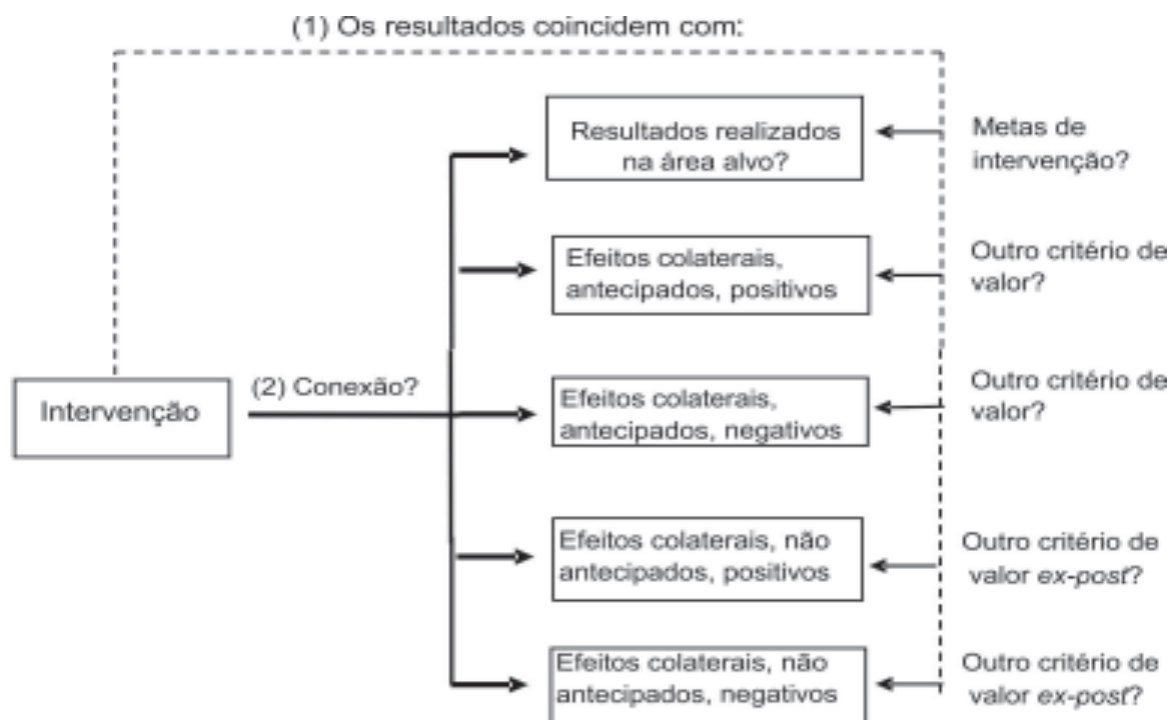
Nesse ponto, é importante apresentar um modelo que, expressamente, considera o importante argumento dos efeitos colaterais, enquanto preserva a fundamental orientação à meta do modelo de alcance de metas: a avaliação de efeitos colaterais.

5.3 Modelo de Efeitos Colaterais

“Há apenas uma regra na política, o inesperado acontecerá.”
Margaret Thatcher

A abordagem de efeitos colaterais é similar à avaliação de alcance de metas no ponto em que os objetivos são mantidos como o critério vital de valor. A novidade com a avaliação de efeitos colaterais é que a procura por resultados almejados é complementada pela busca por subprodutos fora da área alvo. O modelo de efeitos colaterais, assim como o modelo de alcance de metas, é baseado em objetivos, surge daí a expressão “efeito colateral”. Efeitos colaterais ocorrem ao lado dos efeitos principais esperados. Um efeito colateral é um impacto da intervenção que ocorre fora da área alvo. *Um efeito colateral pode ser definido como, pelo menos, uma consequência parcial de uma intervenção, que não pode ser incluída entre os efeitos principais esperados.* O esqueleto básico do modelo de efeitos colaterais é exibido na figura 5.2.

Figura 5.2 – Modelo de Efeitos Colaterais



Efeitos principais são os impactos substantivos centrais que os instigadores da política *intencionalmente* querem atingir. *Os efeitos principais podem ser definidos* como consequências realmente esperadas e desejadas que são, pelo menos em parte, produzidas pela intervenção. Consequentemente, efeitos principais estão associados com os objetivos substantivos da intervenção e com o que se considera possível alcançar. Além disso, efeitos principais são, por definição, antecipados e positivamente valorizados pelos instigadores da política⁷. A ideia subjacente ao modelo de efeitos colaterais é que intervenções públicas podem produzir outras coisas que não os resultados principais. Intervenções podem levar a surpresas positivas mas também criarem novos problemas. Soluções para problemas se transformam em problemas que buscam novas soluções. Soluções para quebra-cabeças se tornam quebra-cabeças que requerem soluções.

O papel das bombas geradoras de calor na produção de energia sueca é um caso ilustrativo de uma solução que se transformou em um problema. A partir da década de 1970 subsídios governamentais foram desembolsados para a instalação de bombas geradoras de calor para recuperação de energia residual como uma solução do problema do petróleo e da energia nuclear. A rápida disseminação de bombas geradoras de calor foi uma consequência de subsídios governamentais. Após alguns anos descobriu-se que o meio de aquecimento, CFCs, nas bombas geradoras de calor poderia vazar para a atmosfera e danificar por muito tempo a camada de ozônio estratosférica que protege a Terra da radiação ultravioleta. Uma pequena parte desses danos pode ser considerado como um efeito colateral não previsto e não intencional do apoio do governo sueco à utilização das bombas geradoras de calor. A partir de meados da década de 1980 bombas geradoras de calor se tornaram um problema ambiental a ser resolvido politicamente.

A questão dos efeitos está resumida na árvore de efeitos na figura 5.3, mostrando quais aspectos podem ser estudados em uma pesquisa de avaliação.

⁷ Alguns usam os termos "efeitos internos e externos, onde aquele resulta diretamente do projeto, por exemplo, eliminação de mosquitos como um efeito interno de um projeto de controle de mosquitos, com abertura de uma área para recreação como um efeito externo" (ANDERSON; BALL, 1978, p. 26).

Figura 5.3
Efeitos Principais, Efeitos Nulos, Efeitos Perversos e Efeitos Colaterais



Resumindo, a avaliação deveria elaborar questões tanto sobre os efeitos principais quanto sobre os efeitos colaterais. Continuando, intervenções públicas podem também gerar *efeitos negativos*. Os efeitos negativos atuam exatamente contra os efeitos desejados ou previstos, por exemplo, pelos instigadores da intervenção. Para utilizar uma terminologia diferente mas pertinente, estes são casos de contrafinalidade. Esses impactos podem ocorrer na área alvo ou em áreas da intervenção pública. Eles também podem ocorrer fora das áreas alvo e serem efeitos colaterais. Os efeitos perversos podem ocorrer muito abaixo na suposta cadeia de governança, por exemplo, em segundo, terceiro ou quarto estágio de resultado (*outcome*). Além disso, eles podem, às vezes, surgir após muitos anos.

O fascínio quase universal sobre o tópico de efeitos negativos indesejados de ações intencionais é manifestado por um conjunto de termos técnicos e históricos – a Astúcia da Razão, a ironia da história, vitória de Pirro, espada de dois gumes, medidas contraproducentes, retrocesso, e efeito bumerangue.⁸

Efeitos perversos são diferentes de *efeitos nulos*. De acordo com Sam D. Sieber, que originou a distinção entre efeitos negativos e efeitos nulos (1981, pp. 10, 47, x), a expressão efeito nulo refere-se à situação onde as intervenções são adotadas porque confia-se e espera-se que ela possa gerar um determinado efeito, mas ele não acontece. As intervenções não produzem impactos nem dentro nem fora de suas áreas alvo. No caso dos *efeitos perversos*, consequências são produzidas, mas são contrárias às pretendidas.⁹

Uma intervenção governamental pode muito bem produzir um efeito nulo e efeitos colaterais ao mesmo tempo. O programa dos reis católicos espanhóis, Fernando e Isabel, de lançar uma expedição marítima liderada por Colombo para encontrar a rota marítima para a Índia pelo Oeste é um caso disso. Colombo não descobriu a rota marítima ocidental para a Índia (efeito nulo na área alvo), mas descobriu a América ao invés disso (efeito colateral fora da área alvo).

Uma vez que, por definição, tanto os efeitos negativos quanto os efeitos nulos ocorrem nas áreas alvo, o modelo de alcance de metas, com toda a sua atenção direcionada ao que acontece nestes domínios específicos, não tem nenhum problema em ser utilizado. Mas isso também quer dizer que o modelo não pode descobrir e verificar efeitos colaterais porque eles surgem fora das áreas alvo.

⁸ Uma ironia histórica não é idêntica a um efeito perverso. Um efeito perverso é, por definição, causado pela intervenção (pelo menos até certo ponto), enquanto uma ironia histórica é um resultado oposto que poderia ter ocorrido independentemente da intervenção. Se alguém diz que um programa levou a um resultado irônico, isso quer dizer apenas que o resultado foi o oposto ao especificado nos objetivos principais e subsidiários do programa. Não precisando ser causado forma alguma pelo programa.

Além disso, é concebível que os objetivos são alcançados sem que sejam provocados pela intervenção. Tal ocorrência de alcance de metas pode ser referida como uma comédia histórica. (Para essa percepção e o termo comédia histórica, reconhece-se o exame de Jesper Schlæger em um seminário realizado na Universidade de Copenhague em 2006).

⁹ Em Boudon (1982, p. 5 et seq.), a expressão “efeitos perversos” é usada, mas no sentido de efeitos não-pretendidos, o que incluiria, é claro, efeitos colaterais.

Efeitos colaterais podem ser *antecipados* e considerados nos cálculos que precedem as decisões para adotar as políticas. Eles podem ser benéficos, assim como prejudiciais.¹⁰

Como consequência da crise do petróleo de 1973, governos por todo o mundo instituíram programas de conservação de energia. Na Suécia, um programa de subsídios governamentais foi adotado para a conservação de energia nas moradias (VEDUNG, 1982a, p. 85). Subsídios estatais foram oferecidos às pessoas que estavam dispostas a melhorarem suas propriedades. O resultado (*outcome*) pretendido era atingir um consumo mais eficiente de energia nos prédios. No entanto, foi antecipado que o programa de subsídios poderia ter um impacto desfavorável na distribuição de renda no país, visto que os proprietários de casas, proprietários de moradias de múltiplas famílias, e outros potenciais beneficiários do programa são geralmente mais ricos que os demais. Embora não fosse cobijado primariamente, os formuladores de políticas públicas estavam preparados para criar e aceitar esse efeito colateral para poder alcançar o objetivo de eficiência energética desejado.

Por outro lado, os apoiadores da política também perceberam que o esquema de subsídios poderia gerar uma série de *spillovers* favoráveis. Impulsionaria as atividades econômicas em geral e criaria vários novos postos de trabalho. Melhorias e isolamentos nas paredes e sótãos diminuiriam a circulação do ar, o que tornaria o clima interior mais confortável. O isolamento e a instalação de janelas de três painéis diminuiriam o barulho.

Um estudo orientado a investigar a conservação energética, assim como a distribuição de renda, emprego e conforto, seria uma avaliação de efeitos colaterais.

Até agora tem sido considerado apenas efeitos colaterais antecipados. Entretanto, alguns efeitos colaterais são, sem dúvida, *inesperados*. Como sua contraparte de efeitos previstos, eles podem ser benéficos ou prejudiciais. O aumento de radiação do gás radônio nas moradias e o aumento da incidência de alergias, prováveis consequências de um melhor isolamento que causa menos circulação do ar, pode ser considerado um exemplo de efeitos colaterais inesperados e prejudiciais.

¹⁰ Efeitos estratégicos causam alguns problemas no discurso sobre efeitos colaterais. O efeito principal desejado com o programa pode ser que o partido no governo vença a próxima eleição. Um possível efeito colateral poderia ser que a oposição se divida. Nesta seção, entretanto, consideram-se apenas os efeitos substantivos.

Consequências benéficas inesperadas são espécimes raros; contudo é provável que os reformadores, tentando vender uma nova política, listem e esgotem todos os possíveis impactos positivos. “O fenômeno de exagerar um programa no momento da formulação e legitimação da política é bem conhecido e, de fato, caracteriza grande parte da elaboração de políticas nos Estados Unidos”, afirmam Ripley e Franklin em um livro sobre implementação (1986, p. 234 et seq.). Entretanto, há ocasiões em que políticas cruéis possuem efeitos colaterais imprevistos e felizes.

A zona desmilitarizada em torno do paralelo 38, formada em 1953 através da península coreana, é uma história interessante de como uma resolução de um conflito armado gerou, inadvertidamente, um último refúgio para espécies ameaçadas de extinção, como o a garça da Manchúria (*Grus japonensis*), o urso nativo coreano, e o gato selvagem. O entorno da zona desmilitarizada de 4000 metros de largura por 238 quilômetros de extensão é uma das áreas mais militarizadas do mundo, assim como é minada, possui obstáculos à movimentação a pé ou de veículos, e é constantemente patrulhada. É também uma das zonas menos habitadas. É o militarismo dessa área que criou, inadvertidamente, um refúgio seguro para a vida selvagem assim como salvou seus habitats das invasões da agricultura intensiva, indústria, construção de estradas e cidades.

A descoberta de que os fenômenos sociais podem ser “o resultado da ação humana, mas não da execução do projeto humano” é geralmente, e erroneamente, creditada a Bernard Mandeville e, particularmente, aos filósofos morais do Iluminismo Escocês como Adam Smith e Adam Ferguson¹¹. Em particular, eles baseiam seus casos para a ideia de mercado livre, irrestrito, na ideia de efeitos colaterais felizes. A famosa formulação de Mandeville “vícios privados, benefícios públicos” refere-se à capacidade do mecanismo do mercado de transformar o egoísmo individual em um benefício coletivo. Os resultados sociais favoráveis acontecem de forma não planejada, como efeitos colaterais do comportamento egoísta dos indivíduos.

¹¹ A memorável frase de Ferguson “resultados da ação humana, mas não da execução do projeto humano” parece ter sido expressa da seguinte forma: “Cada passo e cada movimento da multidão, mesmo nas que se chama de eras iluminadas, são feitos com igual cegueira para o futuro; e nações tropeçam sobre instituições, as quais são de fato resultados da ação humana, mas não a execução de projeto humano. Se Cromwell disse, que um homem nunca monta

O conceito central criado por Adam Smith em seu *A Riqueza das Nações* (1776, 1937, p. 423) foi a “mão invisível”. Em uma passagem ilustrativa, Smith argumenta que cada indivíduo, perseguindo seus próprios interesses, promoveria mais efetivamente, mas sem intenção, o bem comum:

Cada indivíduo necessariamente se esforça por aumentar ao máximo possível a renda anual da sociedade. Geralmente, na realidade, ele não tenciona promover o interesse público nem sabe até que ponto o está promovendo... e orientando sua atividade de tal maneira que sua produção possa ser de maior valor, visa apenas a seu próprio ganho e, neste, como em muitos outros casos, é levado como que por mão invisível a promover um objetivo que não fazia parte de suas intenções... Ao perseguir seus próprios interesses, o indivíduo muitas vezes promove o interesse da sociedade muito mais eficazmente do que quando tenciona realmente promovê-lo.

Um exemplo pode ajudar a esclarecer os diferentes tipos de efeitos estudados até agora. Entre 1974 e 1980, o governo sueco montou um programa para fornecer subsídios financeiros estatais para a indústria, onde o objetivo principal declarado foi o de promover a conservação de energia nos edifícios e processos industriais. Além desse efeito principal, a iniciativa dos subsídios, como pode ser visto na Tabela 5.1, gerou um efeito negativo e efeitos nulos na área alvo, assim como efeitos colaterais positivos e negativos fora da área alvo.

mais alto do que quando ele não sabe onde está indo, então pode ser afirmado com mais razão ainda quanto às comunidades; que elas admitem que, nas maiores revoluções, nenhuma mudança é pretendida, e que os políticos mais refinados nem sempre sabem se estão dirigindo o Estado de acordo com seus projetos” (citado de HAYEK, 1978, p. 264).

Tabela 5.1

Efeitos do programa sueco de subsídios para a conservação de energia em edifícios e processos industriais de 1974 a 1980

1. Efeito principal planejado na área alvo	Redução no consumo de energia
2. Efeitos nulos na área alvo	Diminuição no uso de energia como resultado do apoio dado aos atores que já teriam tomado as medidas (efeito de peso-morto, <i>free riders</i>).
3. Efeitos perversos na área alvo	Aumento do uso de energia a longo prazo por causa da economia de equipamento suplementar ter permitido tecnologias antigas funcionarem por mais tempo.
4. Efeitos colaterais positivos	Ambiente de trabalho aprimorado pela redução das frestas das portas, do chão e das janelas. Redução dos custos de manutenção e operação. Ao preencherem os pedidos de subsídio, os candidatos vieram com outras economias e medidas de eficiência, que foram efetuadas.
5. Efeitos colaterais negativos	Empresas empregam trabalho subsidiado em vez de trabalho não subsidiado (efeito de substituição). As empresas que recebem suporte público superam as empresas que não recebem o apoio (efeito de deslocamento). Destruição de bens de capital porque os equipamentos são trocados mais cedo do que o habitual. Inconveniências para a gestão e para os funcionários em conjunto com a reconstrução e instalação

Sem dúvida, o programa de subsídios trouxe o efeito principal desejado; em muitas ocasiões levou à (1) redução do uso de energia nas indústrias. Em outros casos, a redução no consumo de energia ocorreu como uma consequência das medidas subsidiadas, mas os destinatários que receberam a verba admitiram que essas medidas seriam tomadas mesmo sem o apoio governamental; configurando, (2) efeitos nulos, ou efeitos de peso-morto. Em algumas situações, o programa pode ter levado a efeitos negativos (3) como o aumento do uso de energia a longo prazo; a economia de equipamentos de energia suplementar pode ter mantido tecnologia semi obsoleta operando por um período maior de tempo, impedindo assim uma transição mais rápida para tecnologias mais novas e econômicas que poderiam ter levado a um menor consumo energético a longo prazo. É interessante notar que o programa de subsídios deu origem a efeitos colaterais positivos e negativos. Os três efeitos colaterais positivos (4), assim como os cinco negativos (5) provavelmente não necessitam de nenhuma explicação adicional.

Há um grande debate entre os teóricos sociais sobre contrafinalidade, subprodutos, e consequências não intencionais. Karl Popper, entre outros, em sua obra *Conjectures and Refutations* (1974, p. 124) forçosamente alegou que os efeitos perversos e os efeitos colaterais constituem o problema central para os cientistas sociais:

Uma das coisas mais notáveis da vida social é que *nada vem exatamente da forma que foi planejado*. As coisas sempre saem um pouco diferentes. Nós raramente produzimos na vida social precisamente o efeito que gostaríamos de produzir, e nós geralmente obtemos coisas que não queremos na barganha. É claro que agimos com certos objetivos em mente; mas, além desses objetivos (que podemos ou não alcançar) sempre há certas consequências indesejáveis de nossas ações; e essas consequências geralmente não podem ser eliminadas. Explicar por que elas não podem ser eliminadas é a maior tarefa da teoria social¹².

¹² Albert Hirschman (1991, p. 35) formulou a mesma ideia da seguinte forma: "Uma das maiores percepções da ciência da sociedade—já encontrada em Vico e Mandeville e elaborada durante o Iluminismo Escocês—é a observação de que, por causa das previsões imperfeitas, as ações humanas estão aptas a

A importância de perceber efeitos negativos deve ser óbvia para todos os avaliadores. Mas por que é tão vital prestar atenção aos efeitos colaterais? Por causa dos subprodutos, tanto benéficos quanto prejudiciais, serem fatores cruciais em todo julgamento, abarcando toda a operação de uma intervenção. Caso seja verificado que os efeitos colaterais conhecidos, discutidos e previamente avaliados de forma positiva, não se materializaram após um período de tempo previsto, isso deverá ter consequências para toda a avaliação da intervenção. Se os efeitos no emprego dos subsídios econômicos para construção de isolamentos térmicos são muito mais fracos do que fora calculado antecipadamente, pode-se ter menos razões ainda para manter o programa na forma atual, mesmo que as expectativas sobre a economia de energia tenham se tornado reais.

Os programas podem produzir *spillovers* que se constituem como um problema ou que criam novos problemas que devem ser submetidos a novos programas governamentais. Soluções de quebra-cabeças se transformam em quebra-cabeças que precisam de solução que se tornam um novo quebra-cabeça que requer novas soluções, e assim por diante. Aaron Wildavsky escreve (1979, p. 4, p. 69 et seq.): “Cada vez mais as políticas públicas devem lidar com as consequências das políticas anteriores”. Informações sobre efeitos colaterais são cruciais a qualquer aferição abrangente de programas governamentais.

Efeitos colaterais imprevistos são de particular interesse dos avaliadores. As reais consequências de uma política pública sempre saem, de alguma forma, diferentes do que era esperado pelos agentes antes de entrarem em ação. Isso se deve ao fato de que as consequências são produzidas pela aplicação do programa, e não pelas hipóteses originais que guiaram a decisão a ser adotada.

Até agora estamos longe dos objetivos originais e das intenções subjacentes à intervenção. A ênfase será colocada nos resultados panorâmicos, abrangentes, planejados ou não, previstos ou imprevistos.

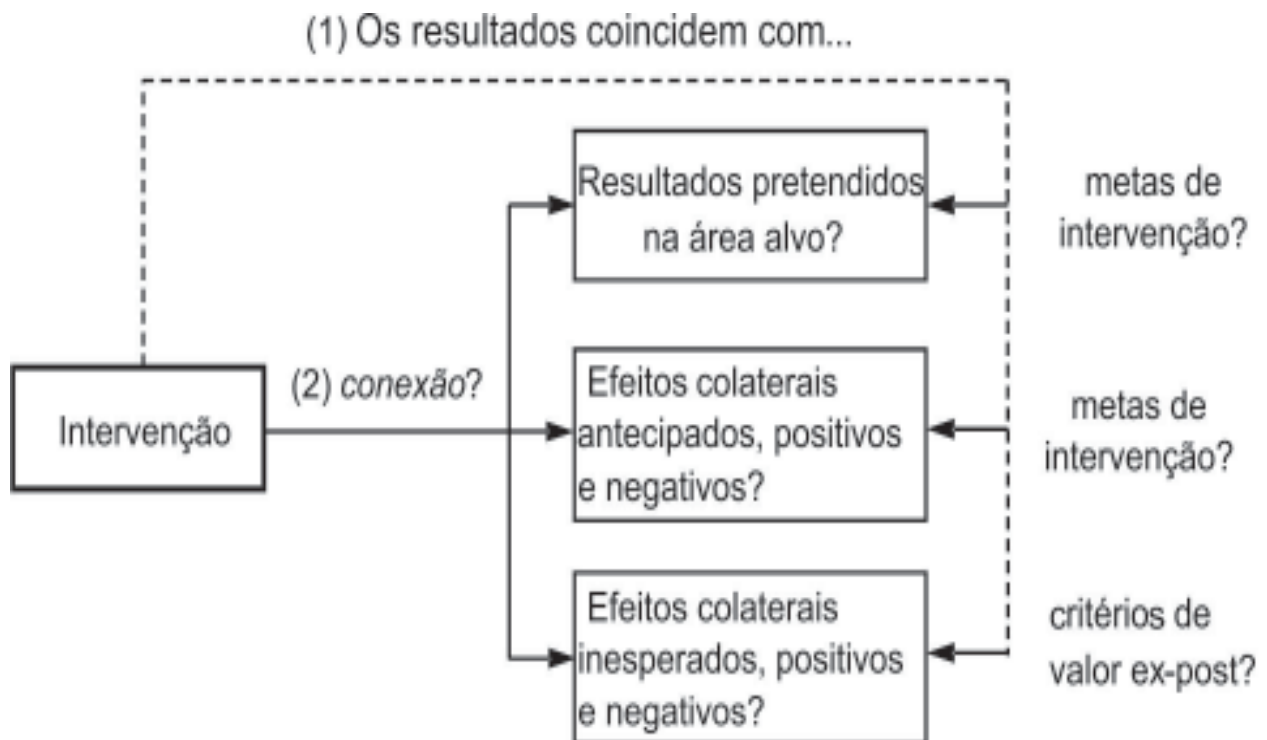
terem consequências não intencionais de escopo considerável. O reconhecimento e a descrição sistemática de tais consequências não intencionais têm sido desde então uma grande tarefa, se não a *raison d'être*, da ciência social.” —Consultar também Boudon (1982), Vernon (1979), Sieber (1981), Hayek (1979, p. 146 et seq), Elster (1978, p. 106 et seq).

Além da teoria das ciências sociais, possíveis sugestões sobre os efeitos colaterais podem ser tiradas da história legislativa da reforma, particularmente de seus oponentes articulados. Outras fontes seriam a erupção de qualquer controvérsia pública alguns anos após a adoção do programa, o que surgiria no governo e no parlamento em conjunção com as reavaliações ou crenças de funcionários pertinentes e outros *stakeholders*, uma vez que o programa já tivesse sido lançado há algum tempo.

Se a totalidade de efeitos (positivos e negativos, consequências antecipadas ou não) de uma intervenção governamental fossem investigadas, a estrutura da avaliação dos efeitos colaterais dos resultados (*outcomes*) poderia ser como mostrado na figura 5.4.

Figura 5.4

Avaliação dos efeitos colaterais com efeitos colaterais especificados



Como na avaliação por alcance de metas, a avaliação de efeitos colaterais é baseada na centralidade democrática da cadeia de comando parlamentar. Diante disto, a abordagem de efeitos colaterais é um forte concorrente do modelo de alcance de metas. De fato, a principal razão para realizar uma avaliação em primeiro lugar é que as ações estatais são imprevisíveis e regularmente resultam em efeitos colaterais involuntários. É uma obrigação da pesquisa avaliativa mapear os efeitos colaterais. “Sempre procurar por efeitos colaterais, mesmo se os objetivos não foram cumpridos” pode ser um mantra para avaliadores independentes. Para concluir: avaliação de efeitos colaterais é uma escolha melhor do que avaliação por alcance de metas.

Um grande desafio é como aferir efeitos colaterais inesperados. Quais critérios devem ser aplicados para julgar seu mérito? De preferência, o avaliador tentaria equilibrar o valor do efeito principal planejado com o valor dos efeitos colaterais benéficos e prejudiciais. Efeitos colaterais negativos podem ser aceitos de má vontade como um custo para alcançar o resultado principal geral. Efeitos colaterais positivos poderiam melhorar a aceitação de programas com poucos objetivos concluídos. O desempenho dessa operação requer critério de valores para efeitos principais, para cada tipo de efeito colateral e uma fórmula para balanceá-los.

A *teoria descritiva de valoração* sugerida pela avaliação por alcance de metas – que declara que os objetivos pré-estabelecidos da intervenção devem ser utilizados como critério de mérito – é claramente insuficiente pelo seguinte motivo: se alguns efeitos não são previstos, o critério para julgar seus méritos e deméritos também não é pré-estabelecido. Portanto, os objetivos são insuficientes para o julgamento individual de cada efeito colateral imprevisto. Pela mesma razão a fórmula para aferir os *trade-offs* entre os valores dos efeitos colaterais e dos efeitos principais também não é pré-estabelecida. Isso tornará impossível para o avaliador calcular o valor global agregado da intervenção usando a teoria descritiva de valoração.

Uma solução viável envolve uma abordagem descritiva diferente e mais criativa para valorar. Além de mapear o efeito principal e compará-lo com os objetivos pré-definidos, ela pode também mapear os efeitos colaterais, mas deixar para os administradores públicos que contratam a avaliação e outros usuários da avaliação determinarem seu valor e realizar a avaliação global do programa *ex-post facto*.

Como o modelo se baseia nos valores de outros, ele ainda abarca uma teoria descritiva de valoração; mas a concentração de efeitos colaterais imprevistos força o avaliador a permitir a valoração descritiva *ex post facto* pelas partes interessadas.

5.4 Avaliação Livre de Metas

Uma resposta drástica para o problema de critério de valor é sugerida através da *avaliação livre de metas*. A avaliação livre de metas foi inicialmente concebida por Michael Scriven (1973, 1974, 1980, 1991).

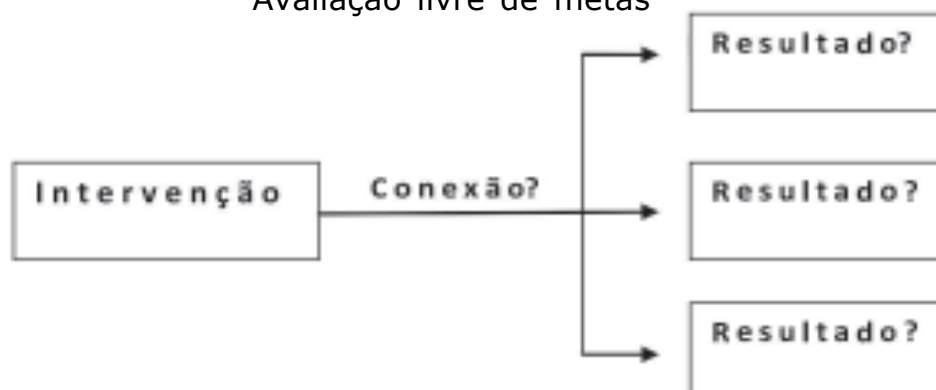
À primeira vista, a contribuição de Scriven parece insolente, quase frívola. Sua insistência de que as avaliações devem ser livres de metas tem que ser um *contradictio in adjecto* (uma contradição em termos). Como alguém poderia avaliar se está proibido de estabelecer metas para aferir? Afinal, o objetivo da avaliação é julgar o mérito, valor e utilidade do que está sendo avaliado. Vários avaliadores têm testemunhado que a sugestão de Scriven, em 1972, sobre o modelo livre de metas foi recebida por uma atordoante descrença. Ignorar completamente as metas e objetivos foi realmente chocante. (GUBA; LINCOLN; 1981, p. 16 et seq.).

O que Scriven está ferozmente reagindo contra é o apego obsessivo dos avaliadores com objetivos premeditados. De acordo com a interpretação da avaliação livre de metas aqui fornecida, o organizador deve ter resultados, sejam planejados ou não. Ao não enviesar a avaliação em direção às metas de intervenção estabelecidas, o avaliador pode ser mais aberto à totalidade dos produtos e resultados do objeto avaliado.

O avaliador sem metas quer concentrar todos os seus esforços em descobrir todos os impactos que a intervenção produziu. O avaliador deve concentrar-se no que o objeto avaliado está fazendo, sem saber o que se está tentando alcançar. Os objetivos premeditados devem não apenas ser desconsiderados; o avaliador deve tomar precauções para não os localizar e compreender seu significado. Ter o conhecimento de metas pré-concebidas e argumentos associados pode se transformar em um espartilho mental, impedindo o avaliador de prestar atenção aos efeitos colaterais, particularmente os efeitos colaterais surpreendentes. A principal tarefa do avaliador é ter uma vista panorâmica da intervenção e descobrir sobre todos os seus efeitos, sem distinguir entre os efeitos intencionais ou não.

Um esqueleto do modelo de avaliação livre de metas é exibido na figura 5.5.

Figura 5.5
Avaliação livre de metas



Avaliação sem metas é, entre outras coisas, uma tentativa de resolver o problema de efeitos colaterais, o cetro que assombra todo planejamento político e avaliação baseada em objetivos. De acordo com Scriven, é uma tentativa de remover as conotações negativas ligadas à descoberta de subprodutos e consequências inesperadas. “Toda a linguagem de “efeito colateral” ou “efeito secundário” ou mesmo “efeito imprevisto” tendiam a ser uma depreciação do que poderia muito bem ser a conquista crucial, especialmente em termos de novas prioridades” (citação em PATTON, 1987, p. 36).

Na própria versão de Scriven do modelo de avaliação sem metas, os efeitos do programa são comparados com as necessidades da população afetada¹³. Nesse ponto, é preciso se afastar da companhia de Scriven, arriscando excluir também as necessidades dos beneficiários do modelo. Nesta reinterpretação conscientemente idiossincrática, o modelo de avaliação sem metas não presta atenção nem às metas pré-estabelecidas, nem às necessidades do cliente. Os problemas envolvidos na determinação e na ponderação das metas do programa ou nas necessidades dos consumidores são desordenadamente desgastantes e dispendiosos de tempo. Deixar de lado o pântano retórico de objetivos e necessidades do avaliador vai poupar tempo precioso que pode ser gasto em tarefas mais urgentes.

¹³ Scriven (1991, p. 180) diz: “Mérito é determinado relacionando os efeitos de um programa às *necessidades* relevantes da população *impactada*, em vez de as *metas* do programa (sejam os objetivos da agência, dos cidadãos, do legislador ou do gerente) para a população-alvo (intencional). Poderia muito bem ser chamado de *avaliação com base nas necessidades* (ou *orientada ao gerente*) em contraste com avaliação baseada no objetivo”.

O modelo livre de metas, na interpretação aqui apresentada, é, aparentemente, relacionado à gestão orientada a resultados (ou gestão por resultados), uma notável técnica de condução do setor público, abordada superficialmente no capítulo três. Na verdade, a *análise de resultados* seria um rótulo adequado. Basicamente, a avaliação livre de metas sugere que os avaliadores devem apenas apresentar os fatos; quanto aos passos adicionais de aplicação de critérios de mérito para os dados realistas, integrando os julgamentos em um valor global e tirando conclusões para a ação futura, é melhor deixar para os tomadores de decisão. Dessa forma, essa noção reformada de avaliação livre de metas aqui apresentada é adequada em contextos de gestão por resultados.

Do ponto de vista da democracia representativa, mesmo o modelo livre de metas reformado sofre de uma limitação. Os objetivos que se quer evitar não são desejos fortuitos ou desejos incidentais. As metas de intervenção adotadas pelos organismos públicos eleitos ou nomeados talvez tenham um estado exultante, uma vez que foram instigadas devidamente pelo povo ou pelos representantes do povo. Deve ser do interesse dos cidadãos que essas metas sejam levadas a sério em avaliações de resultados primários (*outputs*) e resultados secundários (*outcomes*) da intervenção. Essa é uma perspectiva sobre a qual Scriven parece insensível, mesmo 30 anos depois de apresentar pela primeira vez o modelo. Não há na literatura qualquer alusão a esse argumento democrático contra o modelo livre de metas, tanto das contribuições de Scriven quanto das de qualquer outro autor sobre o modelo livre de metas.

O fato de que este modelo livre de metas reformado aparentemente não fornecer critérios pelos quais julgar os méritos dos resultados (*outcomes*) pode ser considerado como uma fraqueza. Uma vez que a valoração deve ser uma propriedade inerente de avaliação, pode-se argumentar que o modelo livre de metas não é absolutamente avaliação. É uma análise dos efeitos ou análise dos resultados, mas não uma avaliação. No entanto, uma rejeição semântica desse tipo é muito limitada. A propriedade de ser livre de metas poderia muito bem se transformar em uma força em um contexto de avaliação: os destinatários da avaliação podem fazer a avaliação. O modelo livre de metas, então, aponta para um acordo institucional do problema do critério de valor na avaliação de políticas públicas. A solução seria: descobrir tudo sobre os resultados do programa, mas deixar a sua avaliação para os tomadores de decisões pertinentes, *stakeholders* e clientes.

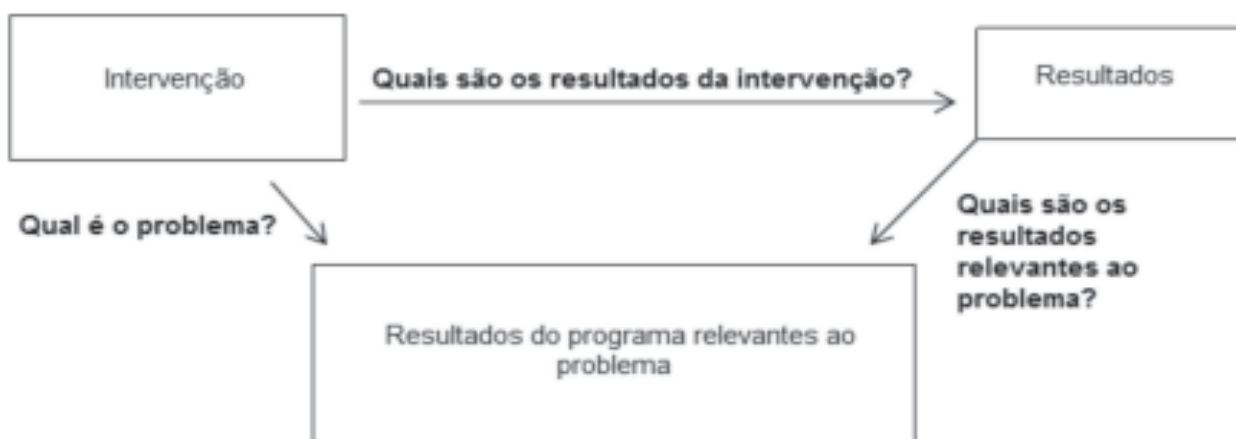
Essa seria uma filosofia que fortaleceria explicitamente a reinterpretação do modelo livre de metas contida neste livro.

Em suma, a avaliação livre de metas ignora completamente objetivos de intervenção e de outros critérios de valor *ex-ante facto*. Isso favorece o estudo puro dos efeitos de resultado. A avaliação livre de metas parece adotar uma filosofia *dê-lhes os fatos*. O papel do avaliador nomeado é ser responsável por resultados em termos de desempenho real e resultados reais, sem prestar atenção às metas iniciais ou outros critérios de mérito. Mas os resultados devem ser comunicados a alguém. Presumindo-se que a seleção, o esclarecimento e a aplicação de critérios e padrões de valor sejam deixados para os potenciais beneficiários *ex-post facto*. Quando dadas as conclusões, os destinatários podem usar quaisquer valores que considerem aplicáveis. No modelo livre de metas, critérios de valor são exercidos apenas após a investigação em si tenha sido realizada.

5.5 Modelos de Relevância

Um déficit presente na avaliação por alcance de metas é que os objetivos podem ser alcançados enquanto problemas subjacentes permanecem. *O modelo de relevância* pretende ser uma solução para isso, porque leva o *problema subjacente* como critério contra o qual o mérito da intervenção é avaliado. Se for relevante para a solução do problema, a intervenção é boa, senão é ruim. *Modelos orientados ao problema* ou *baseados no problema* são outras designações usadas para esta abordagem.

Figura 5.6
O Modelo de Relevância



Problema subjacente é um critério de mérito vago. Qual é o problema subjacente? Aqui, vamos nos referir àquele que é estruturado e inserido na própria intervenção, aquele problema que a intervenção é projetada para atacar. No entanto, é claro que existem outros. Em vez de usar problemas institucionalizados da intervenção, avaliadores podem usar problemas enquadrados e percebidos pelos diversos atores físicos (ou vários atores do tipo ideal), por exemplo, metas, operadores de campo, gerentes de agência, ou os formuladores da intervenção. E não há nada de errado com isso. Aqui, no entanto, não vamos demorar mais falando dessa opção¹⁴.

O que se entende por *relevância*? Caso seja relevante, uma intervenção é “*pertinente* ao assunto em questão”. “*Aplicável* ao assunto em questão” é outra expressão. Esses são os significados do dicionário. Para simplificar também adicionamos *adequado* ao conceito. A intervenção relevante não só é pertinente ou aplicável ao assunto em questão, também é adequada no sentido de ser *apropriada, razoável*, talvez também *satisfatória, suficiente, cabível*.

Uma terceira questão diz respeito ao que é considerado relevante para o problema subjacente. Este texto introdutório vai ficar com as duas respostas fornecidas pelo modelo da União Europeia (UE) na figura 4.3 acima. Efeitos do resultado alcançado na sociedade ou na natureza, ou seja, *relevância do efeito do resultado* são a resposta número 1. Uma intervenção positivamente relevante gerou efeitos na sociedade ou na natureza através da qual o problema subjacente é resolvido ou mitigado ou atenuado; uma intervenção é negativamente relevante se o problema subjacente foi agravado.

Os objetivos da intervenção adotada, ou seja, a relevância das metas, é a resposta número 2. Agora avaliadores examinam se definir metas é relevante para a solução do problema subjacente. Mas isso não é idêntico à avaliação de realização de metas? A resposta é não. Problema subjacente é diferente de meta. A variável *meta* pode não ser idêntica à variável *problema*. Os tomadores de decisão podem ter colocado metas em uma variável da interface da qual se espera a indução de mudanças na variável problema.

¹⁴ Em ajuda para o desenvolvimento um conceito diferente de relevância é utilizado: “ [a]té que ponto uma intervenção de ajuda para o desenvolvimento conforma as necessidades e prioridades de grupos-alvo e as políticas dos países recebedores e dos países doadores (SIDA, 2007).

Pode-se ilustrar com a política pública de redução da destruição da camada de ozônio estratosférica. O chamado buraco negro foi descoberto por cientistas britânicos, que mais tarde viram que o problema indicado por eles foi aceito pela comunidade política planetária. Os objetivos da política global de proteção da camada de ozônio são dirigidos à redução das emissões de certas substâncias químicas, tais como o clorofluorocarboneto (CFC). Não é evidente que o alcance de metas, neste caso, levará à eliminação ou redução do afinamento da camada de ozônio. E isso pode continuar porque a destruição da camada de ozônio é causada por outros fatores além da emissão de CFC.

Aqui terminam os modelos orientados ao problema ou baseados no problema e passa-se a abordar os radicalmente diferentes *modelos orientados ao ator*. Estes surgiram na esteira do aumento da crítica da metodologia empírica positivista e sua fuga dos motivos do ator, intenções, percepções e suposições como ponderações impossíveis de observar e, portanto, de reconstruir empiricamente.

Em modelos do ator, os critérios de mérito são derivados de atores físicos, tais como os seres humanos individuais ou coletivos, como empresas e organizações de interesse. Vamos rever modelos dirigidos ao cliente (público-alvo), as partes interessadas e profissionais.

5.6 Avaliação Orientada aos Clientes (Destinatários)

O modelo orientado ao cliente permite que o público-alvo da intervenção realizem a avaliação, com base em seus próprios critérios de mérito. As questões em jogo são: se a intervenção pública em termos de (1) conteúdo, (2) durante o processo até a sua entrega, (3) no momento da entrega mesmo, (4) no que se refere a obter respostas do cliente ou (5) por meio de seus eventuais efeitos de resultado, alcança a qualidade que os próprios alvos considerem razoável, ou exigem ou querem ver. Como critérios de mérito, o modelo orientado ao cliente pode usar os desejos, pedidos, exigências, objetivos, interesses, expectativas dos participantes, dentre outras coisas¹⁵.

¹⁵ Na verdade, há também uma escola de pensamento que usa *as necessidades* do cliente como ponto de partida para a avaliação orientada ao cliente. Para simplificar, foi evitada a questão das necessidades aqui.

O termo cliente denota os destinatários das intervenções públicas (WEISS, 1991, p. 189, p. 208). Presos em cadeias, alunos de escolas públicas, pacientes em hospitais públicos, idosos que recebem cuidados em asilos municipais, usuários de bibliotecas municipais e passageiros de ônibus e trens são exemplos de clientes (*destinatários, participantes, público-alvo, consumidores, beneficiários, usuários dos serviços*).¹⁶

Várias abordagens diferentes podem ser consideradas orientadas ao cliente. Pense em um exemplo inventado sobre a satisfação do cliente com os serviços do setor público. Na cidade de Helsinki, em 2009, 60% dos pais estavam satisfeitos com o serviço prestado pelas creches municipais. Dois anos depois, o número caiu para 50%. Expressando preocupação com a queda inesperada, o conselho, em janeiro de 2012, decidiu estabelecer 75% de satisfação dos pais como meta a ser alcançada em 2014. Em 2015, um novo estudo demonstrou que em 2014, 65% dos pais estavam satisfeitos. É certo que, tanto o estudo de 2011 como o de 2015 são orientados ao cliente. Esses estudos são avaliações orientadas ao cliente? A resposta é sim. A satisfação do usuário é um critério de mérito orientado ao cliente. Curiosamente, o estudo de 2015 pode ser considerado uma avaliação por alcance de metas, uma vez que 75% de satisfação dos pais foi uma meta estabelecida pelo conselho da cidade.

Nas formas avançadas, os clientes estão muito mais envolvidos. Racionalizando a partir do caso em que a avaliação é (i) encomendada pelos administradores, e (ii) planejada para envolver os usuários de serviços muito mais do que simplesmente pedir-lhes para responder um questionário sobre a sua satisfação com o serviço: incentivam-se os próprios clientes a selecionar os aspectos da intervenção, sua implementação e resultados sobre os quais vão fazer seus próprios julgamentos. Por exemplo, os clientes podem julgar um resultado (*output*) da intervenção, disponibilidade de serviço, qualidade de serviço, ou até mesmo processo de serviço e administração de serviços. O cerne do serviço é projetado para atender às demandas dos clientes? O contato dos clientes com os funcionários do serviço é respeitoso?

¹⁶ A palavra "cliente" ("avaliação centrada no cliente") também pode se referir a quem contrata as avaliações, bem como quem se utiliza dos resultados. Esse não é o significado dado à palavra aqui. (STUFFLEBEAM; SHINKFIELD, 2007, p. 42, p. 212 et seq.). Conforme a *Key Evaluation Checklist* de Scriven (1991, p. 83, p. 85), a palavra "*Cliente*: Quem está contratando a avaliação?" e a palavra "*Consumidor*: quem está usando ou recebendo (os efeitos) do objeto da avaliação?".

Essas são duas perguntas que podem ser respondidas pelos usuários dos serviços em seus relatórios pessoais. Os clientes também podem optar por levantar a questão causal, isto é, estimar os impactos da intervenção estimados sobre si ou sobre a comunidade de clientes em geral¹⁷.

Na *coleta de dados*, os defensores dos modelos orientados ao cliente preferem auto-observação do cliente e um grande número de entrevistas bem conduzidas a questionários, métodos documentais e observação do avaliador. Entrevistar em profundidade alvos individuais é uma técnica muito escolhida, e a distribuição de instrumentos de auto-relatório que os clientes e sua rede pode facilmente completar é outra técnica preferida. Em alguns casos, os avaliadores orientados ao cliente endossam a entrevista do tipo grupo focal (*focus-group*), que permite *deliberações em grupo* entre os participantes e entre os participantes e o avaliador. O avaliador tenta criar fóruns de debate, para promover a riqueza deliberativa. Isso irá apoiar o desenvolvimento de novas ideias, conceitos de serviços, soluções e tecnologias. Ele também pode, como um efeito colateral, educar os participantes, tornando-os melhores cidadãos no futuro.

Atualmente, o modelo de avaliação orientado ao cliente é empregado em vários contextos, tais como casas de repouso para idosos, habitação popular, saúde mental, serviços de utilidade pública, recreação e serviços de saúde física, onde a participação dos clientes é crucial para o funcionamento dos serviços. Modelos orientados ao cliente são utilizados para avaliar as atividades de bibliotecas, teatros, museus, jardins zoológicos e parques nacionais. É um favorito entre os educadores. Nas universidades, os estudantes são rotineiramente convidados a partilhar as suas opiniões de cursos, listas de leitura e palestras. Eles são convidados a avaliar as capacidades de seus docentes para organizar o conteúdo do curso, estimular e promover debates, despertar a motivação e o pensamento crítico dos alunos, e para mostrar preocupação e entusiasmo para os alunos.

¹⁷ Nessa avaliação orientada ao cliente-resultado (*outcome*), o público-alvo tenta determinar a mudança relativa em si mesmo ou no corpo total do cliente, do beneficiário, como resultado de sua participação em modalidades de tratamento específicos. Em termos técnicos, os clientes comparam o que realmente aconteceu com a intervenção/tratamento ao invés de entender o que teria acontecido se não houvesse tratamento ou um tratamento diferente. Avaliação do cliente-resultado usa o *delineamento de controles sombra* na sua avaliação do impacto da intervenção (ver figura 14.1 e capítulo 14).

5.7 Prós e Contras de uma Avaliação Orientada ao Cliente

A intervenção de *ajuste aos desejos do cliente* é uma grande justificativa para o modelo orientado ao cliente. Alguns filósofos alegam que a pressão dos clientes expressa através de atitudes e sugestões de melhoria como resultado e de avaliações levará à melhoria não só do serviço principal, mas também de processos de serviços, prestação de serviços, efeitos do serviço e satisfação do cliente. Através da avaliação orientada ao cliente, serviços públicos tornam-se mais claramente voltados para os desejos e expectativas dos clientes/fregueses.

No entanto, a avaliação orientada ao cliente não pode ser levada muito adiante, uma vez que a noção do cliente inclui aspectos participativos e deliberativos que podem funcionar como *educação para a cidadania democrática*. A característica participativa sugere que os clientes também são cidadãos que podem exprimir as suas queixas e desejos para os avaliadores, gestores e prestadores de serviços, deliberar entre si e com outros *stakeholders*, e, em certa medida, influenciar e assumir a responsabilidade pelo conteúdo do serviço. O recurso deliberativo engendra um permissivo diálogo discursivo, racional e de aprendizagem que pode educar os prestadores de serviços a prestar atenção às preocupações dos clientes e clientes a se tornarem melhores cidadãos: o consumidor como cidadão e não o consumidor como cliente¹⁸.

Em terceiro lugar, a avaliação orientada ao cliente reforça a *legitimidade da intervenção*. Se os clientes são convidados a dar opiniões e a ter alguma influência sobre a intervenção em formação, processos e resultados, a sua aceitação do sistema provavelmente vai aumentar. Além disso, a avaliação orientada ao cliente pode promover a *eficácia e eficiência*, porque a concentração nos clientes pode forçar prestadores de serviços e gestores a acabar com muitas preocupações além de prestar bons serviços.

¹⁸ A grande experiência de Evert Vedung como professor de avaliação de políticas públicas o ensinou a enfatizar a diferença entre *clientes da intervenção* (participantes da intervenção, destinatários) e *clientes da avaliação* (usuários da avaliação). O cliente da avaliação é a pessoa, grupo ou agência que solicitou a avaliação de uma intervenção ou aquele que deve utilizar seus resultados, enquanto o cliente da intervenção é o pretendido ou real receptor da intervenção.

Mas os avaliadores devem estar cientes da tendência da clientela a exagerar as queixas, a fim de obter mais serviço. Os clientes também podem nutrir *ilusões fiscais*. Maior envolvimento do cliente na avaliação pode entregar o poder a grupos com interesses pessoais limitados.

O modelo impulsionado ao cliente pode complementar os modelos de realização de metas e efeitos colaterais, uma vez que é muitas vezes relacionado com processos e resultados da implementação, enquanto realização de metas e os efeitos colaterais são focados em efeitos nos resultados. Além disso, a exclamação do proprietário do hotel suíço Cesar Ritz "*le cliente n'a jamais tort*" ("o cliente nunca está errado") não pode ser a estrela-guia de toda a avaliação. A exigência da capacidade de resposta do serviço público às preocupações do cliente é boa, mas dentro de certos limites. Ela nunca pode ter precedência sobre o requisito de que os operadores da linha de frente devem seguir as diretrizes de seus superiores administrativos e, indiretamente, dos órgãos políticos, como o Parlamento, o Conselho Municipal, e, em última análise, dos cidadãos como uma coletividade, cujos votos determinaram a composição e orientação política geral desses corpos. Os idosos em uma comunidade que aproveitam o auxílio social municipal domiciliar não podem tomar decisões que vão de encontro às regras de agentes e dirigentes na cadeia representativa de controle. Eles não podem unilateralmente diminuir taxas de serviço, por exemplo. Autoridades fiscais devem responder pronta e corretamente a perguntas de contribuintes, mas elas não podem cumprir a demanda dos contribuintes para impostos mais baixos ou dar isenções ao acaso. Modelos de avaliação fundamentadas na democracia representativa devem ter precedência sobre modelos orientados ao cliente.

5.8 Avaliação do Stakeholder

A *avaliação do stakeholder* procura refletir a pluralidade de perspectivas sobre qualquer intervenção num setor público. De acordo com Majone (1989, p. 183), a ideia não é tanto fornecer medidas objetivas e aferições de resultados "quanto é facilitar um diálogo amplo entre os defensores de diferentes critérios". Essa é uma abordagem racional comunicativa de avaliação, na tradição de Jürgen Habermas de *situação de discurso ideal e racionalidade argumentativa*, não a racionalidade meios-fins de Weber.

Na avaliação do *stakeholder*, preocupações e assuntos de vários atores afetados servem como critério de mérito quando intervenções – particularmente sua organização, processos administrativos, resultados primários (*outputs*) e conteúdos e, mais raro, seus resultados secundários (*outcomes*) – são aferidos e avaliados. *Stakeholders* podem ser *definidos* como grupos ou atores individuais que são afetados por ou tem algum interesse na intervenção a ser avaliada, na sua execução ou nos efeitos de seus resultados. Interesse pode ser medido em termos de dinheiro, status, poder, aparência, oportunidade ou outra moeda, e pode ser grande ou pequena, construída pelos grupos em questão (GUBA; LINCOLN 1989, p. 51). Isso é bem diferente da utilização de objetivos pré-fixados como critério de mérito, como na avaliação de realização de metas. Avaliação do *stakeholder*, no entanto, se assemelha ao modelo orientado ao cliente, sendo a maior diferença o escopo: enquanto o modelo impulsionado ao cliente está preocupado com uma categoria de interesses afetados, o modelo do *stakeholder* é orientado a todas. A figura 5.6 mostra um levantamento de potenciais *stakeholders* em uma intervenção de bem-estar social local (GUBA; LINCOLN 1989, p. 40; RIECKEN; BORUCH 1974, p. 203, ROSSI; FREEMAN 1989, p. 422; WEISS, 1972b, p. 18). Um conceito bastante semelhante à avaliação dos *stakeholders* é a avaliação participativa.

Figura 5.7
Potenciais Stakeholders em Intervenções Públicas Locais



A avaliação do *stakeholder* acontece de diferentes maneiras. Os *stakeholders* podem constituir eles mesmos o time de avaliação e executá-la. A avaliação pode também ser conduzida por avaliadores particulares, que se reúnem com os *stakeholders* para identificar seus pontos de vista. Em ambos casos, o modelo é participativo. Na sequência, examina-se aqui o caso em que a avaliação do *stakeholder* seja feita por avaliadores particulares.

A avaliação, tendo como base o *stakeholder*, começa com o avaliador mapeando os grupos envolvidos mais importantes, afetados por ou que tem algum interesse na emergência, adoção, execução e resultados da intervenção. O avaliador identifica as pessoas que iniciaram, batalharam, financiaram e adotaram a intervenção, i.e. basicamente os criadores da intervenção. Ele identifica aqueles que estão encarregados da implementação: gestores sênior, intermediário e júnior, funcionários, e operadores da linha de frente, que são os que realmente entregam o resultado da intervenção. Ele destaca o grupo-alvo primário da intervenção, os clientes e suas associações. Ele identifica parentes e suas associações. Ele pode também incluir leigos. E ele procura por *stakeholders* latentes que sabem que têm interesses na intervenção, mas preferem manter um perfil discreto, com aqueles que desconhecem seus interesses, e completamente passivo.

Uma vez que os maiores interessados são discernidos, Egon Guba and Yvonna Lincoln, em seu livro *Effective Evaluation* (1981, p. 33 et seq.; também 1989, p. 50 et seq.), propõem que as “preocupações” e “questões” dos *stakeholders* devem ser o ponto de partida. Uma *preocupação* é “qualquer assunto de interesse ou importância para uma ou mais partes”. Pode ser algo que os ameaça, algo que eles temem que possa levar a consequências indesejáveis para eles, ou algo que estão ansiosos para fundamentar. “Praticamente qualquer reclamação, dúvida, medo, dificuldade prevista e similares expressas por qualquer pessoa com uma base legítima para fazer tal representação pode ser entendido como uma preocupação”. Uma *questão*, por outro lado, “é qualquer afirmação, proposição, ou foco que permite a apresentação de diferentes pontos de vista; qualquer proposição sobre a qual pessoas razoáveis podem discordar, ou qualquer ponto de contenda”.

Defensores da avaliação do *stakeholder* nutrem uma forte inclinação por métodos qualitativos e coleta de dados interativa. Avaliadores devem falar com *stakeholders* para extrair suas histórias e dados observacionais, que por sua vez devem ser permitidos a afetar o próximo passo do avaliador no processo de pesquisa. Depois de um tempo, ele pode descobrir ambos objetivos do programa, falsos e genuínos, e quais as preocupações que os *stakeholders* nutrem com relação ao programa. Com o tempo, o avaliador fica mais envolvido e pode começar a determinar quais dimensões e preocupações dos *stakeholders* devem ser incluídas no estudo.

É típico para os modelos do *stakeholder* que o avaliador possa procurar extensivamente pelas dimensões e pelas preocupações cruciais nelas presente. A ideia é que o avaliador tem que responder às questões e preocupações dos atores afetados e deixar que essas governem o próximo passo nessa iniciativa investigatória. Através de comunicação interativa, ele deve descobrir quais dimensões e preocupações dos *stakeholders* devem ser levadas a sério e sondadas mais profundamente. A concepção da avaliação é *emergente*, ou seja, gradualmente determinada. A avaliação dos *stakeholders* é uma *avaliação responsiva* (STAKE, 1975; SHADISH Jr; COOK; LEVITON, 1991; p. 275 et seq.).

Para obter os dados finais, os defensores do modelo dos *stakeholders* apresentam as mesmas preferências como as do modelo do cliente em relação aos métodos qualitativos, interativos e dialógicos (veja acima).

Depois que os dados são acumulados e processados, o relato dos resultados, o que pode variar de um para outro *stakeholder*, terá início. A palavra-chave parece ser 'representações', isto é, caracterizações ricas em informação com imagens, anedotas, descrições densas e citações. A visão holística abrangente mediada através de uma representação é importante. Normalmente, serão incluídos vários critérios de mérito, padrões de desempenho nesses critérios e várias avaliações abrangentes. Na prática, o modelo do *stakeholder* se tornará absolutamente pluralista. Finalmente, é decidido se as descobertas devem ser colocadas em um relatório escrito, algo que Guba e Lincoln (1981, p. 25 et seq., p. 39 et seq.) e outros não consideram de forma alguma necessário.

5.9 Prós e Contras da Avaliação de *Stakeholder*

Avaliação de *stakeholder* possui várias *vantagens*, das quais quatro serão mencionadas aqui.

- 1 O argumento democrático
- 2 O argumento do conhecimento
- 3 O argumento da utilização
- 4 O argumento do compromisso

O(s) *argumento(s) democrático(s)* começa(m) dos pontos de vista participativo e deliberativo. Na verdade, democracia significa que, nas eleições, cidadãos votem em elites concorrentes que deveriam tomar decisões em seu nome (*democracia representativa*). Porém, os cidadãos deveriam também poder participar na decisão pública final entre eleições (*democracia participativa*). Além disso, discussão, diálogo, debate, argumentação, são também importantes valores democráticos, pois ajudam pessoas a formar e a refinar suas crenças e preferências (*democracia deliberativa*). O modelo do *stakeholder* satisfaz esses valores – participação e deliberação – mais do que os modelos de realização de metas e efeitos colaterais. Pode-se dizer que o modelo do *stakeholder* traz uma *abordagem argumentativa para avaliação*, não uma abordagem científica, como pode ser o caso do modelo de realização de objetivos.

De acordo com o *argumento do conhecimento*, *stakeholders* nutrem convicções sobre efeitos colaterais inadvertidos, barreiras de implementação e absoluta trapaça, o que pode fornecer ao avaliador ideias sobre os tópicos para uma investigação mais aprofundada. Já que a orientação ao *stakeholder* trará mais aspectos do assunto em discussão, a qualidade das descobertas da avaliação irá aumentar.

Na verdade, toda avaliação deveria começar com a determinação de quem são os atores relevantes e rodadas de coleta de dados interativos, a fim de ter uma rápida e provisória noção do significado, implementação e resultado da intervenção. Honestamente, um aspecto impressionante do modelo do *stakeholder* é que esse evoluiu de um modelo singular de avaliação independente para uma situação, onde vários elementos foram integrados em outros modelos e abordagens, também em combinação com a abordagem da teoria da intervenção. (DONALDSON, 2007, p. 10 et seq., p. 33 et seq., p. 39, p. 40 et seq., etc).

A abordagem do *stakeholder* aumenta as chances de abordar as questões de interesse genuíno das partes envolvidas. Ela traz à luz informações que atendem às reais necessidades dos diferentes *stakeholders*, aumentando assim a probabilidade de que os resultados de fato sejam colocados em uso. Esse é o *argumento da utilização* em apoio à avaliação dos *stakeholders*.

Finalmente, o modelo do *stakeholder* promove *compromissos* e previne conflitos políticos. Assembleias de *stakeholders* são mecanismos de construção de consenso. Eles são veículos para moldar um acordo sobre os resultados de esforços anteriores, e mais importante, formular as propostas de ação futura. A nomeação de um grupo de avaliação do *stakeholder* marca uma tentativa de chegar a um acordo entre os diversos públicos dissidentes. Construção de consensos e dar legitimidade às decisões fundamentais são consideradas grandes vantagens da avaliação do *stakeholder*.

Há também *desvantagens* óbvias nos modelos de *stakeholder*. Avaliações dos *stakeholders* são *desmedidamente impraticáveis* e *exigem muitos recursos*, uma vez que cada público específico de cada *stakeholder* deve ser contatado e ouvido com atenção.

Modelos de *stakeholders* são *nebulosos (fuzzy)*. Nenhuma resposta oficial é dada para a questão de quem são os *stakeholders*. A gama de interessados deve ser decidida caso a caso. Além disso, todos os públicos de *stakeholders*, seja lá como foram selecionados, são tratados como iguais. Mas, em um sistema democrático constitucional, políticos eleitos devem ter um peso maior do que os administradores ou especialistas sobre os assuntos substantivos considerados. O modelo do *stakeholder* não incorpora nenhuma prioridade entre os *stakeholders*. Como a avaliação orientada ao cliente, a avaliação do *stakeholder* deve trabalhar dentro das estruturas fixadas pela democracia representativa.

Há uma tendência, no entanto, no atual discurso e prática de considerar a participação dos *stakeholders* não só como um modelo específico, mas também como uma abordagem participativa e argumentativa que deve ser usada, do início ao fim, na maioria das avaliações: nesse sentido, o envolvimento do *stakeholder* é muito parecido com a abordagem da teoria de intervenção.

5.10 Modelos de Avaliação por Colegiado: Crítica dos Pares, Autoavaliação e Combinações

Modelos de avaliação por colegiado são particularmente aplicados em campos dominados por um complicado e informal critério de mérito e profissões fortes. Modelos de colegiado diferem muito de modelos baseados em metas e modelos econômicos em tomar os critérios de qualidade pertinentes à profissão como ponto de partida da avaliação. Dessa forma, os advogados avaliam advogados; cientistas, cientistas; cirurgiões, cirurgiões; enfermeiros, enfermeiros; educadores acadêmicos, educadores acadêmicos; e assim por diante. Outro termo para a avaliação de colegiado é *revisão pelos pares*.

A avaliação por colegiado é a aferição do trabalho por pessoas de competência semelhante aos produtores do trabalho (colegas, iguais), à luz de seus critérios de mérito profissional ou vocacional em comum e de seus padrões de qualidade.

A avaliação por colegiado vem em diferentes disfarces. A mais célebre é a *revisão cega pelos pares*, usada para aferir artigos acadêmicos individuais e ensaios para jornais ou conferências acadêmicas. Nesse caso, os nomes dos autores não são dados aos avaliadores e o nome dos avaliadores não são conhecidos pelos autores. Em casos ideais, os revisores não sabem a identidade do autor, já que qualquer informação de identificação é removida do documento antes da revisão. Além disso, as duas categorias devem ser mantidas isoladas uma da outra. Não deve haver nenhuma comunicação pessoal direta entre autores e avaliadores. O sistema foi projetado para reduzir ou eliminar a parcialidade. É a chamada dupla avaliação cega pelos pares. O editor do jornal ou organizador da conferência, por outro lado, sabe os nomes dos indivíduos de ambas categorias¹⁹.

Instituições de pesquisa e de ensino superior são julgadas através de outro tipo de avaliação por colegiado, baseada na franqueza, não no sigilo, e onde os avaliados conhecem os avaliadores e vice-versa. Além disso, a revisão por pares é combinada com autoavaliação pelo avaliado. O procedimento geralmente começa com a autoavaliação. O departamento, programa ou projeto a ser avaliado pede a seus profissionais para produzir um relatório do seu próprio desempenho.

¹⁹ http://en.wikipedia.org/wiki/Anonymous_peer_review#Anonymous_peer_review (acesso 05/12/2013).

Eles discutem os pontos fortes, pontos fracos, oportunidades e ameaças (usando o chamado *esquema-SWOT*). Os resultados são agregados ao nível da instituição.

Em seguida, os cientistas externos de renome e educadores do campo particular são designados para avaliar a qualidade e relevância do desempenho da instituição. Não só esses colegas baseiam a sua aferição preliminar sobre a autoavaliação produzida pelos avaliados; eles também fazem visitas nas instituições pertinentes para recolher informações adicionais a partir de apresentações orais dos avaliados, conversas com avaliados e observações. Avaliadores e avaliados se envolvem em diálogos, discussões, palestras, conversas. Em suma, eles se comunicam uns com os outros. Uma vez que os colegas compilam um relatório preliminar, a comunicação continua, já que os avaliados são convidados a apresentar as suas observações escritas sobre o relatório preliminar, e é esperado que os avaliadores leiam e reflitam sobre essas observações quando finalizarem seu relatório. Ao fim, avaliadores e avaliados leem os rascunhos escritos, conversam entre si várias vezes, e solicitam opiniões um do outro. Eles se envolvem em processos argumentativos dialógicos muito diferentes da ideia de sigilo embutido em revisão cega por pares.

É possível se perguntar o que modelos de colegiado têm a ver com a avaliação do setor público. Esses modelos parecem estar quilômetros de distância do conteúdo de intervenção, implementação, produção e resultados. A resposta está incorporada no princípio do setor público orientado à profissão. Em algumas áreas da vida pública, as metas são tão complexas e as técnicas de governança tão difíceis que autores da Constituição (os constituintes originários) acharam mais sensato deixar a formação e debate delas para profissionais com uma boa formação. Arquitetos, juízes, professores, médicos, veterinários e engenheiros seriam casos em que isso se aplica. Por isso, também é considerado natural delegar a avaliação para as profissões. Mas, uma vez que esses profissionais trabalham no setor público, revisão por pares deve ser considerada como um modelo de avaliação em pé de igualdade com os outros modelos utilizados na vida pública.

Será ilustrado esse protótipo de avaliação com a antiga estratégia da Agência Nacional Sueca para a Educação Superior para avaliação de pesquisa e educação nas universidades e faculdades do país. A Agência uma vez instituiu uma política de avaliação nos seguintes 10 passos:

1) *Política de avaliação de agência.* O Conselho adota uma política compreensiva: durante os próximos seis anos, todas as universidades e faculdades no país serão avaliadas numa base rotativa.

2) *Agrupamentos de avaliações.* O Conselho decide em termos gerais sobre os itens específicos das avaliações: por exemplo, próximo ano, a pesquisa em todos os departamentos acadêmicos de serviço social será avaliada.

3) *Pares externos + gestor interno do projeto.* Para cada avaliação particular, o Conselho aponta um painel de pares externos qualificados (colegas, especialistas) que recebem informação introdutória e um gestor de projeto (projeto em grupo) dentro da agência.

4) *Autoavaliações por avaliandos, avaliados.* Cada departamento de serviço social (objeto da avaliação) produz uma autoavaliação a partir dos requerimentos da Agência. Itens da avaliação devem ser postos sob três enunciados: (i) pré-condições para a pesquisa, (ii) processos da pesquisa e (iii) resultados da pesquisa. Os resultados das autoavaliações são classificados no esquema SWOT, i.e., Forças, Fraquezas, Oportunidades e Ameaças (em inglês).

5) *Pares externos consultam as autoavaliações.* Os especialistas externos se familiarizam com a questão, consultando autoavaliações, relatórios de pesquisas, CVs e outros materiais relevantes dos que estão sendo avaliados.

6) *Visita ao local pelos pares.* Os especialistas externos executam suas aferições através de *visitas*. No geral, os que estão sendo avaliados, pesquisadores e grupos de pesquisas fazem apresentações de seu trabalho e participam de discussões cara-a-cara com os colegas externos. Os colegas observam, escutam, fazem perguntas, discutem e fazem anotações.

7) *Pares externos compilam aferição preliminar.* Os especialistas externos compilam aferições preliminares de cada departamento a ser avaliado.

8) *Pares externos escrevem o relatório final.* O relatório final é preparado pelos especialistas externos e o gerente do projeto formula uma aferição final, que será a base onde o Conselho decidirá sobre medidas como o cancelamento da concessão de funcionamento de cursos ou o direito a dar exames específicos em disciplinas.

9) *Conferências de follow-up para disseminação de resultados.* Conferências de follow-up são organizadas, onde cada departamento é informado sobre as aferições particulares feitas pelos revisores externos.

10) *Follow-up pelo conselho da agência.* Um ano depois, o Conselho faz um follow-up para verificar se as medidas recomendadas estão sendo tomadas.

Uma característica central da avaliação por colegiado é que o critério de mérito e os padrões são escolhidos pelos próprios profissionais, pois o conhecimento especializado é necessário, a fim de entender e aplicá-los. Ademais, os critérios são complexos por estarem associados a metodologias específicas e às questões de cada disciplina. Além disso, esses recursos estão mudando continuamente. Acrescente ainda que os critérios de qualidade são muitas vezes comunicados oralmente apenas nas profissões relevantes como o *conhecimento tácito*. Esse conhecimento tácito deve ser trazido à tona através de diálogos esclarecedores entre grupos de colegas.

Uma segunda força é que as visitas são curtas e, portanto, de baixo custo. Por outro lado, a avaliação ainda pode tornar-se cara por honorários de consultores, despesas de viagem, e porque é preciso tempo para concluir o relatório final.

Um ponto fraco, com curtas visitas ao local e autoavaliações é que os que estão sendo avaliados podem ser tentados a mostrar fachadas e erguer aldeias Potemkin (ver nota de rodapé no capítulo 7).

Modelos de avaliação por colegiados frequentemente produzem resultados instáveis. Painéis pareados têm demonstrado usar critérios de mérito e padrões de desempenho amplamente diferentes e chegar a conclusões diversas. No entanto, em áreas tecnicamente complexas, a avaliação por colegiado é provavelmente o melhor método disponível para julgar a qualidade do que é produzido²⁰.

²⁰ A organização modelo genérico de avaliação pelos pares de uma instituição de pesquisa universitária, o roteiro passo-a-passo e as forças e fraquezas é fornecida lucidamente por House (1993, p. 56-7)

MODELOS ECONÔMICOS

Comum a todos os modelos unicamente de substância é o foco singular nos resultados substantivos em vez de nos custos. Por mais dispendioso, em termos de recursos, que o programa possa ter sido, avaliadores unicamente de substância se concentram nos resultados substantivos (e nos processos que precedem esses resultados) e negligenciam os custos. Por outro lado, atenção a ambos —substância e custo— é um aspecto típico dos modelos de avaliação econômica.

Modelos de avaliação econômica são modelos de avaliação que levam em conta custos nos seus resultados. O valor econômico da eventual realização de metas por unidade monetária gasta é contraposto à eficácia e aos benefícios oriundos da realização para aferir a eficiência da intervenção realizada.

6.1 Avaliação Econômica Inclui Custos de Intervenção

Toda intervenção implica em custos. O desenvolvimento de políticas, programas e outras iniciativas criam custos, pois alternativas precisam ser sondadas, investigações técnicas precisam ser feitas e relatórios precisam ser produzidos. Depois da adoção, a intervenção deve ser administrada, o que requer operadores, gerentes, executivos, espaço físico, material de escritório e equipamento técnico. Executivos de agências, gerentes, funcionários da linha de frente e destinatários devem ser informados sobre a existência e conteúdo de novas políticas e programas. Folhetos informativos e prospectos devem ser produzidos e distribuídos, e jornalistas e publicitários contratados e treinados para disseminar as informações. Regulamentações como legislação habilitante requer pessoal com *know-how* técnico que possa analisar os pedidos. Depois dos pedidos serem aprovados, necessita-se de pessoal para supervisionar, avaliar e receber informações sobre o feedback. O mesmo é verdade para instrumentos de política econômica e informação. Empréstimos e subsídios são distribuídos mediante solicitações, o que requer pessoal para o processamento dos pedidos e tomadas de decisão. Eles também

pressupõem vigilância pós-fato, revisão de desempenho e informação sobre o feedback. Os destinatários das intervenções também implicarão em custos quando preencherem os formulários, participarem nas negociações etc.

Modelos econômicos são comumente divididos em unicamente de custo, produtividade e eficiência. O modelo mais simples é o unicamente de custo, singularmente focado no aumento ou redução de custo. Terá a intervenção custado menos este ano do que no ano passado? Terá sido a intervenção implementada do modo mais barato possível? Como os custos podem ser reduzidos? Em que casos podem os gastos crescentes ser controlados e reduzidos? Este modelo unicamente de custo não será aqui analisado, concentrando a análise nos de produtividade e eficiência.

6.2 Avaliação Econômica: Produtividade

Frequentemente, negócios privados são postos como um ideal a ser emulado pelo setor público. Num tipo ideal de negócios privados, lucro é o critério de mérito e maximização do lucro o padrão de sucesso. Em analogia a isso, maximização da produtividade deve ser o padrão de um bom desempenho para intervenções públicas. Produtividade parece ser um conceito estranhamente claro. É o relacionamento entre saída de produtos e a entrada de recursos para que essa saída seja produzida. Em resumo: *output* (saída) dividido pela entrada (custo)¹.

Produtividade pode ser expressa através de um simples algoritmo mostrado na figura 6.1.

Figura 6.1
Produtividade

$$\text{Produtividade} = \frac{[\text{valor do}] \text{ output (saída)}}{[\text{valor da}] \text{ entrada}}$$

¹ No entanto, Stone (2012, p. 63) parece tratar produtividade por seu termo eficiência.

A razão da figura 6.1 pode ser operacionalizada de várias maneiras. Usando um exemplo artificial de uma biblioteca comunitária, computa-se a produtividade das bibliotecas municipais com as seguintes medidas:

$$\frac{\text{Número de livros emprestados}}{\text{Custo em termos monetários, p. ex. Euros}} \quad (1) \text{ Custo da Produtividade}$$

Alternativamente, os avaliadores recorrem à produtividade do trabalho, o que pode ser ilustrado pela expressão:

$$\frac{\text{Número de livros emprestados}}{\text{Custo em termos físicos, por exemplo, número de horas trabalhadas.}} \quad (1) \text{ Produtividade do trabalho}$$

A diferença é que os custos no caso anterior são indicados em termos monetários, e no segundo caso em horas trabalhadas, isto é, em entidades físicas. A unidade de tempo usada pode ser o ano fiscal, ano calendário, ou até um período mensal.

Aqui estão mais alguns exemplos de métricas para produtividade.

Controle de poluição da água	Litros tratados por 1000 euros gastos
Imposto de renda	Imposto coletado por 1000 euros gastos
Processamento de passaporte	Nº de pedidos processados por pessoa-ano
Processamento de passaporte	Unidade de custo em euro por pedido
Parques	Custo em euro por hectare mantido
Áreas de piscina pública	Número de visitantes por 1000 euros
Inspeção de normas de construção	Pessoa-hora ou custo euro por inspeção
Combate a incêndios, prevenção	Pessoa-hora ou custo euro per capita
Combate a incêndios, prevenção	Respostas por pessoa-hora ou por 1000 euros
Trânsito urbano	Viagem de passageiros por 1000 euros

Para dizer algo que vale a pena sobre a verdadeira produtividade de uma agência pública, um caso de referência é necessário. Para uma visão geral, consulte a seção 6.4 Pontos de Referência abaixo.

Produtividade como uma métrica carrega vantagens técnicas. Às vezes, custos não são terrivelmente difíceis de calcular já que eles chegam até agências como fundo monetizado. Contudo, dividir os recursos financeiros em custos propriamente relevantes para cada projeto em particular pode, às vezes, tornar-se uma tarefa árdua, pois os fundos concedidos, alocados e gastos são geralmente indicados como soma total para o programa como um todo.

Porém esses obstáculos parecem pequenos em comparação com as dificuldades de se medir resultados substantivos, apesar de produtividade somente pressupor que são indicados em termos físicos, e não monetários.

Voltemos ao caso da biblioteca municipal, para ilustrar a dificuldade de encontrar indicadores válidos de *output* (saída). Permitir que o público em geral tome livros emprestados é sensatamente a tarefa mais importante de uma biblioteca pública. Mas será o empréstimo de livros um indicador exaustivo do *output* de uma biblioteca? De jeito nenhum. Só 30%-50% dos usuários da biblioteca pegam livros emprestados, segundo um relatório finlandês. Os outros leem jornais, revistas e periódicos. Eles frequentam o departamento de música para escutar ou pegar emprestado gravações, fitas e discos. Eles assistem a exposições ou palestras (SJÖBLOM, 1991; p. 18 et seq.). Tudo isso será ignorado se o avaliador utilizar livros emprestados como indicador de *output* da biblioteca. Na investigação finlandesa a que se refere, esses problemas foram percebidos e discutidos antes de os pesquisadores decidirem escolher pelo número de livros emprestados como indicador de *output* da biblioteca.

Outro problema é que a produtividade não mede qualidade. Há diferenças de qualidade entre livros de autores ganhadores do prêmio Nobel, como José Saramago, Gabriel Garcia Marques e Elias Canetti, e de autores populares como Jackie Collins e Jean Marie Auel (nascida Jean Marie Untinen).

Há outras fraquezas. O mais importante é que produtividade é uma métrica interna, a qual não apreende o que realmente se quer desembaraçar, nominalmente: os efeitos que a intervenção, sua implementação e resultados produziram nos receptores ou na sociedade como todo, o valor desses efeitos e se seus benefícios fazem os custos valerem a pena. No exemplo da biblioteca, livros

emprestados não é uma medida firme; pessoas frequentemente pegam livros emprestado da biblioteca, os colocam numa pilha na mesa de casa, e após algumas semanas os devolvem intocados. Mais importante e decisivo é a leitura dos livros emprestados. Mas o que realmente importa são os ganhos advindos da leitura de quem pega emprestado, seja uma leitura educacional, seja recreativa. O que o avaliador consciente sobre o custo, orientado à educação, realmente quer compreender pode ser:

Valor da educação e recreação através de livros emprestados e lidos

Custo da produção desse valor

Mas se assim for, então ele terá deixado a medição da produtividade e terá entrado no campo dos efeitos dos resultados e da avaliação de eficiência.

Produtividade não é um parâmetro ideal para aferir o valor de instituições do setor público. Produtividade mede quantidades, enquanto qualidades tendem a ser negligenciadas. E a instituição pública pode fazer coisas erradas, i.e., os *outputs* (primeiros resultados) podem não resultar nos *outcomes* (segundos resultados) desejados.

6.3 Avaliação Econômica: Eficiência

Eficiência é tida como o valor da administração pública. Em seu trabalho clássico *Administrative Behavior* (1976, p. 186), o vencedor do prêmio Nobel Herbert Simon argumentou este ponto: "O critério que o administrador aplica aos problemas factuais é um de eficiência. Os recursos, a entrada, à disposição do administrador são estritamente limitados... É sua função maximizar o alcance dos objetivos governamentais (assumindo que eles já foram acordados) pelo emprego eficiente dos limitados recursos que estão disponíveis para tal". Eficiência tem sido estimada também em outros contextos. George Bernard Shaw (1907, ato 4): "Há somente duas qualidades no mundo: eficiência e ineficiência, e somente dois tipos de pessoas: a eficiente e a ineficiente."

Enquanto produtividade mede *output*/custo, eficiência mede *outcome*/custo. Eficiência é expressa como a razão do valor do *outcome* (resultado) produzido pela intervenção sobre os custos da intervenção. Por exemplo, o número de usuários de serviço tratados por pessoa-ano, número de usuários de serviço satisfeitos tratado por pessoa-ano ou simplesmente *outcome* dividido pelos custos. Ou o valor do *outcome* criado por euro gasto.

Antes de prosseguir, uma ressalva deve ser feita a respeito de “eficiência” e “eficácia” e “efetividade”. Na aferição da *eficiência*, custos são incluídos, enquanto a aferição da *eficácia* se concentra nos resultados substantivos sem levar em consideração os custos. Eficácia é realização de metas e eficiência é alcance de metas por euro. Neste livro, o uso da eficácia da intervenção não tem nada relacionado a custos.²

Eficiência pode ser medida de dois jeitos, como custo-benefício ou como custo-efetividade. “Aferições de eficiência (análises de custo-benefício e custo-efetividade) fornecem um quadro de referência para custos relacionados para resultados de programas”, escreveram Rossi e Freeman em seu amplamente utilizado livro *Evaluation*: “Em análises *custo-benefício*, ambos *entradas (input)* e *resultados (outcome)* do programa são medidos em termos monetários; nas análises custo-efetividade, *entradas (input)* são estimadas em termos monetários e *resultados (outcome)* em termos do verdadeiro impacto” (1989, p. 375).

Se mensurada numa análise custo-benefício, eficiência pode ser expressa como a razão do valor monetizado do *outcome* produzido pela intervenção sobre os custos monetizados. Se equalizada para o que é medido em uma análise de custo-eficácia, eficiência presta atenção aos custos monetizados como na análise de custo-efetividade, mas o valor dos efeitos é indicado apenas em termos físicos. Isso está expresso nos algoritmos na figura 6.2.

² **Eficiência**: é quando algo é realizado da melhor maneira possível, ou seja, com menos desperdício ou em menor tempo. **Eficácia**: é quando um projeto/produto/pessoa atinge o objetivo ou a meta. **Efetividade**: é a capacidade de fazer uma coisa (**eficácia**) da melhor maneira possível (**eficiência**).

Figura 6.2 - Eficiência como (1) Custo-Benefício e (2) Custo-Efetividade

<i>(1) Eficiência (custo-benefício)</i>	=	$\frac{\text{valor dos efeitos da intervenção (monetizados em Euros)}}{\text{custo da intervenção (monetizados em Euros)}}$
<i>(2) Eficiência (custo-efetividade)</i>	=	$\frac{\text{valor dos efeitos da intervenção em termos físicos}}{\text{custo da intervenção (monetizados em Euros)}}$

“Efeitos da intervenção” na figura 6.2 não são idênticos a nenhuma ocorrência na área selecionada depois de iniciada a intervenção. Algumas dessas ocorrências podem ter sido causadas por outra coisa. O que avaliadores devem procurar são efeitos, no mínimo parcialmente e indiretamente, produzidos pela intervenção, ponto final. Assim, aferição da eficiência usa a mesma medida que o modelo de alcance de metas. A grande diferença é que avaliação da eficiência leva custos em consideração, o que a análise de alcance de metas não faz.

6.4 Pontos de Referência

Para ter uma impressão de quão boa a produtividade medida (ou eficiência) de uma intervenção realmente é, ela deve ser comparada a algo. Tecnicamente falando, avaliadores precisam de *padrões de desempenho* nos critérios de produtividade para dizer algo sensato sobre a verdadeira produtividade. O cientista político israelense Yehezkel Dror (1968, p. 28) apresentou uma visão geral do padrão de valor mais comumente utilizado em avaliação. Com algumas mudanças, a lista de padrões de performance de Dror segue abaixo. Para simplificar, usou-se produtividade como uma métrica.

1. O passado. Como a produtividade obtida é comparada com a do passado? Quanto o empréstimo de um livro custava em 2015 num município A comparado com 2014, 2013, 2012, ou ainda o preço médio num período longo de, digamos, 2004-2014?

2. Comparação intranacional. Como a produtividade alcançada é comparada com a de instituições similares numa mesma área municipal, regional ou nacional? Quanto o empréstimo de um livro custava na biblioteca L em comparação com a biblioteca M, N e O presentes no mesmo município A? Ou quanto custava o empréstimo de um livro em 2015 no município A comparado a bibliotecas similares em municípios B, C ou D no mesmo país?³

3. Comparação internacional. Como a produtividade alcançada é comparada com as de instituições similares de outros países? Quanto o empréstimo de um livro custava em 2015 nos municípios finlandeses A, B, C e D comparado a medidas similares nos municípios suecos A, B, C e D?

4. Ponto de referência. Como a produtividade alcançada em 2015 se sai em comparação com a melhor prática empírica (ponto de referência, *benchmark*)?

5. Mínimo. É a produtividade alcançada alta o suficiente para cumprir as demandas mínimas?

6. Ideal. É a produtividade alcançada a melhor possível de acordo com um modelo ideal?

Assumindo que medidas de produtividade e eficiência possam ser desenvolvidas e adotadas, os benefícios parecerão extensos. Primeiro, conscientização dos resultados em si estimula as pessoas a melhorar. Indivíduos tendem a mirar mais alto quando veem os efeitos de seus empreendimentos. Se há um contínuo *feedback* baseado em alguns critérios de resultados objetivos, pessoas serão motivadas a melhorar suas contagens. Isso, por sua vez, impulsionará a satisfação do funcionário e o moral.

Provisão da produtividade e eficiência medidas pode também tirar a ênfase de diretivas, normas e regulamentos e até da supervisão. Trabalhando com um bom conjunto de medidas de produtividade e eficiência, o funcionário pode se tornar seu próprio mestre. Critérios objetivos guiam o trabalho de um empregado e determinam seus sucessos melhor do que um supervisor.

³ Cf. (SHADISH Jr; COOK; LEVITON, 1991, p. 88).

Todos esses fatores podem promover inovação; uma vez postos, os objetivos específicos tendem a se tornar mais importantes do que os procedimentos pelos quais procura alcançar. Outro ponto forte, mas duvidoso, é a habilidade do modelo de comprimir o valor de uma atividade em um simples número, fácil de compreender e de lembrar.

O perigo com todos esses modelos de avaliação – e particularmente com os modelos econômicos devido a sua aparente precisão matemática e concisão – é que eles podem ser aplicados de modo não crítico e que tomadores de decisão e outros *stakeholders* podem erroneamente acreditar, ou fingir acreditar, que um modelo pode dispor respostas compreensivas e finais.

Assim, é importante manter em mente que todo modelo oferece somente perspectivas e respostas parciais. Por essa razão, triangulações de vários modelos são recomendadas.

6.5 Sobre Valor e Critério de Mérito

É tarefa do leitor (ao digerir o relatório) “avaliar”, no sentido literal do conceito, e é tarefa do avaliador fornecer ao leitor as informações que ele pode querer tomar em consideração na formação de seu julgamento.

Lee J. Cronbach

Ruim é ruim e bom é bom, e é o trabalho da avaliação decidir qual é qual.

Michael Scriven

Os últimos dois capítulos ofereceram uma abrangente visão geral de nove modelos de avaliação. Os modelos foram classificados na base da *dimensão do objeto* (os aspectos da intervenção avaliada e dos processos que a acompanham, tratados nos modelos) e da *dimensão de valor* (o critério de mérito usado pelos modelos para aferir o valor da intervenção). Para finalizar este capítulo, far-se-á ainda uma discussão a mais sobre os assuntos de valoração e critério de mérito.

Em valoração, o precioso é separado do sem valor, como se separa o joio do trigo. Partindo de critérios de mérito e padrões apropriados, aferições deveriam deixar claro se intervenções são inestimáveis, apreciadas, aceitas, duvidosas ou fracassos absolutos.

A atividade de valoração pode ser dividida em dois grandes problemas e cinco subproblemas.

Figura 6.3

Oito Problemas em Avaliação de Políticas Públicas (OPA) - Problema # 7

7 Problema do valor

7.1 Considerações de princípios. Valoração descritiva ou prescritiva?

7.2 Considerações de princípios. Antes-do-fato, durante-o-fato ou após-o-fato? Quando o critério de avaliação deve ser determinado?

7.3 Considerações de princípios. Por qual critério específico de valor e padrões de desempenho nesses critérios de valor deve o valor da avaliação ser julgado?

7.4. Aplicação prática. Quais são os reais valores do objeto avaliado em cada critério e padrão?

7.5. Aplicação prática. Qual é o valor real global ou o valor composto ponderado do objeto avaliado?

Observação: ver capítulo 7 a seguir.

As subquestões 7.1-7.3 tratam opções teóricas de princípios, enquanto que as subquestões 7.4 e 7.5 de aplicações práticas de alternativas das opções de princípios nos objetos sendo avaliados (*evaluands*) concretamente. Nas seções subsequentes abaixo, as opções 7.1 a 7.4 serão um pouco mais esclarecidas. Algumas opções com relação a aplicações práticas serão também consideradas.

Valoração descritiva ou prescritiva? Seguindo os passos de Shadish Jr., Cook e Leviton em seu formidável tratado *Foundations of Program Evaluation* (1991, p. 46 et seq.), valoração descritiva ou prescritiva são as opções na primeira situação de escolha. *Na valoração descritiva*, o avaliador escolhe os valores dos outros como critérios e padrões. Esclarecendo, ele faz uma estimacão de como a intervenção ocorre à luz de suas próprias metas ou, alternativamente, na visão dos clientes, operadores ou *stakeholders* da intervenção em geral.

Na valoração prescritiva, o próprio avaliador “advoga a primazia de valores específicos”, como, por exemplo, justiça, igualdade, igualdade de gênero ou necessidades dos clientes, não importando se esses valores estão incorporados na intervenção ou adotados por qualquer corpo tomador de decisões ou assegurado por algum público de *stakeholder*. Teorias prescritivas de valoração mantêm alguns valores como superiores a outros enquanto teorias descritivas representam valores de outros sem contestá-los ou alegar que um valor é melhor que outro.

Teorias prescritivas dão aos avaliadores uma perspectiva externa crítica que não se compara com as teorias descritivas. Elas abrangem as percepções dos avaliadores do que é uma boa intervenção ao expandir seus conhecimentos do que é benéfico para a condição humana em geral.

Valoração prescritiva encontra dificuldades porque não há uma teoria do valor prescritivo correto. É preciso que o avaliador se posicione nas questões que valem a pena. É claro que eles podem fazê-lo e argumentar sua escolha com cuidado e abertamente. Em contraste, ao aceitar o critério de mérito de outros, o avaliador pode manter sua objetividade naquele assunto e ficar longe da problemática opção de adotar uma postura pessoal na questão dos valores.

Valoração descritiva é mais fácil de lidar com do que a prescritiva, já que tudo o que o avaliador tem que fazer é perguntar a outros. Valoração prescritiva é intelectualmente mais exigente e requer treinamento em ética.

Teorias descritivas são mais compatíveis do que as prescritivas com organizações sociais e políticas de democracias do Ocidente, como Suécia, França e Canadá, baseadas na manutenção do pluralismo de valores que competem um contra o outro em arenas políticas e administrativas.

Adotando a valoração descritiva, avaliadores aumentam as chances de que suas descobertas sejam vistas como legítimas e úteis. Em contrapartida, advogar uma ética prescritiva e acumular dados nessa base, não refletirá nem o pluralismo nem os valores adotados pelas intervenções, e a probabilidade de que tal informação seja percebida como legítima é menor, fazendo que seja menos crível na política. Teorias descritivas são mais legítimas e, portanto, mais úteis que teorias prescritivas nesse sentido (SHADISH Jr.; COOK; LEVITON, 1991, p. 47 et seq.).

Como Shadish Jr., Cook e Leviton, sugere-se que ambas aferições, prescritiva e descritiva, são necessárias no setor público. Valoração prescritiva é melhor realizada externamente por instituições de ensino superior, pesquisa básica e aplicada, pela grande mídia e outros atores críticos e independentes. Valoração descritiva é convencional e, assim, politicamente e socialmente mais prática em sistemas de interesses pluralistas. A maioria dos avaliadores adotam a valoração descritiva. Eles retratam valores declarados nos documentos da intervenção ou assegurados por *stakeholders*, aplicam esses critérios aos dados sobre o desempenho, apresentam a avaliação dos diferentes componentes separados uns dos outros e deixam a integração total a cargo dos potenciais usuários da informação.

Antes, durante ou depois do fato? Os critérios de mérito de uma avaliação, devem ser determinados: 1) antes do fato (previamente, i.e. medidas pré-determinadas, pré-ordenadas ou pré-estabelecidas), 2) durante o fato (enquanto o processo de avaliação está se desdobrando), ou 3) depois do fato (uma vez que a coleta avaliativa de dados, a análise destes e o resumo dos dados estiverem completos)? Essas são as opções da segunda escolha fundamental com respeito ao problema da valoração (7.2 na figura 6.3). A resposta clássica “previamente” vem do modelo de alcance de metas e metas pré-determinadas. Já que o modelo do ator se tornou cada vez mais popular, a resposta tem sido mais e mais “durante” ou “depois do fato”.

Critério específico? Uma vez que as escolhas descritiva/prescritiva e antes/durante/depois do fato sejam tomadas, os atores envolvidos devem determinar os critérios específicos de mérito. Nessa visão geral dos modelos presentes nos capítulos 5 e 6, os critérios de mérito específicos exibidos na figura 6.4 já foram esmiuçados:

Figura 6.4
Critério de Valoração nos Nove Modelos de Avaliação

- 1 Metas de Intervenção (antes do fato)**
- 2 Metas de Intervenção + critério depois do fato para efeitos colaterais inesperados (antes do fato + depois do fato)**
- 3 Sem meta alguma - modelo livre de metas (somente critério depois do fato)**
- 4 Problema subjacente (antes do fato)**
- 5 Preocupações do público-alvo, expectativas e concepções de qualidade (durante)**
- 6 Critérios de mérito de diversos públicos de *stakeholders* (modelo do *stakeholder*)**
- 7 Critério de qualidade de pares externos e avaliadores (antes do fato, durante)**
- 8 Produtividade, *output* : custo (antes do fato)**
- 9 Critério de eficiência, efeitos no *outcome*: custo (antes do fato)**

Uma variedade de outros critérios viáveis , em adição às nove opções abrangidas nos modelos, é apresentada na figura 6.5.

Figura 6.5
Critérios de Valoração Adicionais para
a Aferição de Intervenções Públicas

-
- | |
|--|
| <ol style="list-style-type: none">1 Igualdade (tratamento justo)?2 Legalidade?3 Tratamento digno dos clientes da intervenção?4 Publicidade e transparência?5 Legitimidade (aceitação como correto e apropriado)?6 Representatividade?7 Igualdade de gênero, integração de imigrantes, acessibilidade para pessoas com deficiência?8 Inovatividade?9 Processos internos férteis?10 Sinergia - adicionalidade?11 Simbolismo?12 Sustentabilidade?13 Democracia (poliarquia)? |
|--|
-

Uma norma processual profundamente enraizada na administração pública ocidental é a *igualdade legal* (POLLITT, 2003, p. 9; HECKSCHER, 1958; ECKHOFF, 1989, p. 242). Girando em torno do gerenciamento de casos, a igualdade exige que a pessoa em causa seja tratada de forma justa pelas autoridades pertinentes. Deve ser permitido que obtenha documentos e outros materiais preparatórios que a agência pretende usar e ser oferecido a ela a chance de argumentar a seu favor antes de ser tomada uma decisão formal. Depois, o indivíduo deve receber uma justificativa para a decisão e ser notificado sobre como apelar para uma autoridade superior. Ademais, casos iguais devem ser tratados de forma igual. Por exemplo, o servidor público responsável não deve fazer favores para parentes, colegas do partido, amigos da família ou pessoas da mesma universidade, nem desfavores para inimigos, pessoas nascidas em uma determinada província especial (no sistema unitário sueco) e adversários políticos.

Em *Bureaucracy*, Wilson (1989, p. 126 et seq.) usou os programas de aquisição da defesa americana para ilustrar o caso de igualdade legal e justiça. O Departamento de Defesa, através da Agência de Logística de Defesa, compra todo ano comida, combustível, roupas e peças de reposição que custam bilhões de dólares, e administra quase 200 bilhões de dólares em contratos governamentais. O Congresso e o presidente repetidamente deixaram claro querer o sistema funcionando de maneira eficiente. Há, contudo, restrições legais da justiça colocadas no processo de aquisição para garantir que os contratos sejam adjudicados de forma justa, i.e., assim permitindo que todos os interessados possam competir em condições de igualdade. “As regras essenciais são que deve ser oferecido a todos os potenciais adjudicatários oportunidade igual para fazer uma oferta no contrato; que a decisão de aquisição da agência deve ser justificável objetivamente na base de especificações por escrito; que os contratos adjudicados na base de propostas fechadas, devem ir para o adjudicatário que oferece o menor preço; e aos licitantes que não forem aceitos deve ser oferecida a chance de protestar das decisões das quais discordam”.

Caracteristicamente, os requisitos de equidade tornam a gestão de caso mais demorada. Equidade processual é cara. Critérios orientados aos resultados como eficiência, eficácia ou produtividade conflitarão

com a equidade processual e outros critérios para que os processos sejam corretos, justos. *Trade-offs* entre igualdade processual de um lado e eficiência, eficácia ou produtividade de outro, não podem ser realizadas através de um cálculo técnico-científico.

Outros valores processuais são *legalidade* – compatibilidade com leis, regulamentos, normas oficiais, convenções existentes - *tratamento digno aos clientes da intervenção e publicidade e transparência* nos processos administrativos. *Legitimidade* é “o valor onde algo ou alguém é reconhecido e aceito como correto e apropriado”. Na Ciência Política, *legitimidade* é comumente compreendida como aceitação popular e reconhecimento, pelo público, dos processos e resultados de intervenções públicas ou de um regime de governo em geral (ROTHSTEIN, 1994, p. 120 et seq.).

A consequência tácita pretendida de parcerias locais entre as autoridades municipais, associações da sociedade civil, empresas e cidadãos leigos, que solicitaram e receberam o apoio dos Fundos Estruturais da União Europeia, pode ser moldar o reconhecimento, a legitimidade das propostas e atividades que as parcerias propõem. Já que aqueles que serão afetados são convidados para desenvolver e propor planos de ação incluindo objetivos e medidas, eles também são considerados dispostos a assumir responsabilidade, defender e implementar esses planos quando adotados. Por fazer estruturas de rede e processos legitimarem o processo subsequente, resultados também ganharão legitimidade. Incidentalmente, a ideia é provavelmente aumentar também a legitimidade geral da UE entre cidadãos de países membros.

Outra medida de aferição possível é *representatividade*. Em um serviço público representativo, agências possuem uma força de trabalho que reflete a diversidade da população nacional como um todo. A administração é um microcosmo de uma sociedade maior, por exemplo, com relação a gênero, línguas, região e raça. O serviço público sul-coreano, por exemplo, tem sido avaliado do ponto de vista do equilíbrio regional nos padrões de recrutamento. Esses estudos descobriram que pessoas das províncias de Kyongsang são mais representadas em comparação com as pessoas das províncias de Cholla. Enquanto se aumenta a legitimidade, a representatividade social no serviço público pode ou não ser tecnicamente a mais competente e eficiente.

A União Europeia pode fornecer outro exemplo de *representatividade*. Além da integração de imigrantes e acessibilidade para pessoas com deficiência, a UE recomenda seus avaliadores a utilizar igualdade de gênero como o chamado *valor horizontal*. Programas e projetos financiados pela União e dirigidos ao desenvolvimento de regiões desfavorecidas, redução da exclusão social e outras metas verticais deveriam considerar esses três valores horizontais também.

Inovatividade é outro indicador. Ocasionalmente, espera-se que a cooperação em projetos resulte em algo diferente e incomum. Numa gestão baseada em resultados, os agentes de implementação são conferidos com alguma liberdade de ação para escolher meios e métodos eficientes para alcançar metas estabelecidas por seus superiores, um arranjo em que se acredita que liberará criatividade e criará inovação. Inovação significa "a qualidade de ser novo: frescor, modernidade, novidade, hodierno, originalidade". E é exatamente o que é antecipado aqui.

De acordo com a perspectiva do *processo interno fértil*, uma rede colaborativa é valiosa se sua vida interna é caracterizada por elevada frequência dos agentes implementadores, os argumentos perspicazes, sérias trocas de significados, reuniões curtas, mas intensas e decisões proveitosas no agrupamento ou na troca de ativos. Aparentemente há uma forte sobreposição entre essa perspectiva e o indicador de inovatividade.

O critério do processo interno fértil aponta para os indicadores de vários aspectos de procedimentos internos e funcionamento de reuniões de projeto, tais como frequência e regularidade, mas também frequência do *stakeholders* nas reuniões. Será que os representantes das organizações participantes vão sempre às reuniões ou eles não aparecem? Quem é enviado como substituto? Eles deixam as reuniões antes de elas acabarem? As sessões são sempre presenciais ou são realizadas também por Skype, telefone, e-mail ou em algum outro formato menos demorado? São apontados gestores de processos específicos? Como eles trabalham? Processos internos também incluem preparação do ator, preocupações do ator sobre o assunto tratado e atividade do ator. Entusiasmo no trabalho é crucial. Se o entusiasmo reina entre os parceiros colaboradores, o projeto será provavelmente frutífero e criativo. Tensões e conflitos são benéficos, mas também

podem ser destrutivos. É o trabalho caracterizado por uma deliberação abnegada, uma negociação honesta e acomodação mútua? E quanto a confiança entre os participantes? Esses são exemplos de questões que podem iluminar e julgar os procedimentos de acordo com o critério do processo interno frutífero.

"*Sinergia* é a interação de múltiplos elementos num sistema para produzir um efeito diferente ou maior do que a soma dos efeitos individuais. O termo *sinergia* vem do grego *synergia* de *synergos*, significando 'trabalhar junto'. No contexto do comportamento organizacional, seguindo a visão de que um grupo coeso é mais do que a soma de suas partes, sinergia é a habilidade de um grupo de ser melhor do que seu melhor membro." (<http://en.wikipedia.org/wiki/Synergy>, acesso 29/11/2013). Outro termo para sinergia é valor adicionado.

Sinergia ocorre tipicamente de dois jeitos. Imagine que os atores 1, 2 e 3 perceberam que eles possuem uma meta em comum, que não pode ser alcançada se cada um agir por si só, porém é capaz de ser alcançada se unirem recursos e agir em conjunto. Sinergia aparece se eles colocarem suas vantagens juntas e assim realizar mais do que realizariam se cada pessoa buscasse a meta individualmente.

No segundo caso, os atores podem trocar recursos um com o outro sem se unir num só grupo. Imagine que as agências A, B e C decidiram cooperar. A agência A ganha algo da agência B, o que possibilita que aquela alcance suas metas; em troca ela dá algum recurso para B, que A sente que pode dispor, mas que permite que B alcance suas metas em alguma outra área. Então A pede a C por algo que C está disposto a dar, assegurando que esta ganhará algo em troca de A no futuro que permitirá que C realize suas metas na área que melhor atua. Através dessas trocas colaborativas, todos os três atores participantes estarão melhores em comparação a uma situação sem colaboração alguma.⁴

⁴ Wikipedia oferece dois exemplos. 1) "pessoa A sozinha é muito baixa para alcançar a maçã na árvore e a pessoa B é pequena também. Uma vez que a pessoa B senta nos ombros da pessoa A, elas são altas o suficiente para alcançar a maçã. Nesse exemplo, o produto da sinergia das duas seria uma maçã." 2) "Outro caso seria dois políticos. Se cada um consegue reunir 1 milhão de votos sozinhos, mas juntos conseguem atrair 2,5 milhões de eleitores, sua sinergia teria produzido 500 mil votos a mais do que se cada um trabalhasse independentemente." (<http://en.wikipedia.org/wiki/Synergy>, acesso em 04/12/2005).

De acordo com a perspectiva do *simbolismo*, uma intervenção é valiosa na medida em que é assim percebida por *stakeholders* internos e externos como menciona Hanne Foss Hansen (1989). Não tem que funcionar bem tanto internamente quanto externamente; o ponto é que seja percebida com bom funcionamento e performance pelos públicos-chave tanto internos quanto externos. Se a intervenção e seu sistema de implementação têm a capacidade de sinalizar seu valor, então é valiosa.

Em assuntos ambientais, de agricultura, manejo florestal etc., *sustentabilidade* significa “não levar à exaustão dos recursos ou degradação do meio ambiente”. Em outras áreas, carrega conotações mais gerais como “capaz de ser mantido ou defendido” ou “capaz de se manter em uma certa taxa ou nível”. O segundo significado ocorre frequentemente no desenvolvimento de política de auxílio. Já que o apoio do país doador ao mundo em desenvolvimento vem na forma de projetos, autoridades doadoras querem saber se os resultados alcançados serão mantidos quando o projeto acabar e o financiamento do projeto cessar. Esse é o significado do critério de sustentabilidade no auxílio para o desenvolvimento.

E, por último, mas não menos importante, *a democracia* oferece um conjunto de critérios que podem ser utilizados na valoração da intervenção pública. Para esclarecer, propõe-se um exemplo de critério de avaliação de uma sociedade política inteira. Quando se avalia “o status democrático da União Europeia”, Christer Karlsson (2001) deduziu cinco amplos critérios de democracia a partir da definição de Robert Dahl de um processo completamente democrático em sua discussão sobre *poliarquia e democracia* (DAHL, 1989, p. 221 et seq., em especial a Tabela 15.1, p. 222). Ademais, Karlsson operacionalizou esses critérios em onze indicadores. Apesar de não discutido nesta pequena introdução, seu completo conjunto de critérios está aqui exibido, para uma visão geral, na figura 6.6.⁵

⁵ Democracia participativa e deliberativa como critérios de avaliação são extensivamente discutidos em Vedung (2009) e (2006).

Figura 6.6

Critério de Avaliação para um Processo Completamente Democrático

Critério de valor	Indicadores
I Votação—Igualdade	1. Representação igual 2. Prestação de contas
II Participação	3. Resposta de intermediários 4. Eficácia de intermediários 5. Distribuição de acesso a intermediários
III Deliberação	6. Acesso à informação 7. Existência de uma esfera pública
IV Controle de Agenda	8. Grau de controle de agenda indireta 9. Distribuição de controle de agenda indireta 10. Possibilidade de recuperar autoridade delegada
V Inclusão	11. Condições de cidadania

Fonte: Adaptado de Karlsson (2001, p. 43, p. 44-103).

Contagem global. A alocação de um valor único no desempenho geral ou no caráter do objeto sendo avaliado (avaliando) é o dever supremo da avaliação. Poucos, mas não muitos, teóricos de avaliação têm argumentado a favor dessa posição a respeito do critério real de avaliação. Por exemplo, eficiência seria combinada com os valores de justiça processual, de representatividade e de legitimidade em um único valor geral. Essa ambição, no entanto, não pode nunca ser realizada por meios científicos. Na análise final, a *contagem global* precisa de juízo político. Um jeito de sair desse dilema é apresentar a avaliação de cada um dos fenômenos mencionados separadamente e deixar a contagem global para os tomadores de decisão, como uma questão política.

Nos anos 1960 e de novo nos anos 2000, houve uma tendência tecnocrática na pesquisa de avaliação. Jogos políticos míopes e mal concebidos devem ser suplantados no longo prazo, com base científica, e tomada de decisão racional. Mas a avaliação não pode nem substituir totalmente, nem deve ser permitido substituir o julgamento político. Em todas as alocações autoritativas (sob controle de uma autoridade) de valores, para usar a famosa definição de David Easton sobre o sistema político (1965), há difíceis problemas de equilíbrio, que podem ser resolvidos através da política, e não através de cálculos técnicos, por mais que sejam confiáveis e realizados cuidadosamente.

Deixe-me, sobre este assunto, citar Lennart Lundquist (1990, p. 18):

Em conflitos tanto dentro quanto entre diferentes tipos de critérios de mérito, não há medida à mão que, de forma objetiva, estabeleça quais valores devam ser aplicados. Tampouco existe sempre um sistema de sanções que garante que todos os atores cumprem um possível critério dado. Como são definidas as prioridades nesses casos? Como uma regra geral, o problema não pode ser resolvido através da aplicação de técnicas, mesmo as mais (tecnicamente) sofisticadas. O único jeito remanescente é decidir através do poder. Isso significa que alguém ou algo (por exemplo, servidores públicos, políticos eleitos, partes numa negociação ou o mercado) com poder para fazer passar a sua vontade, é quem, em última análise, decide quais critérios e normas deveriam ser aplicados. Isso mostra a óbvia natureza política da avaliação.

Concorda-se, neste livro, fortemente com Lundquist em sua defesa da política contra afirmações exageradas da pesquisa de avaliação. Avaliação, mesmo que científica e sistemática, não pode, nem deve, ser uma substituta para o alcance do desempenho político – controvérsia, discussão e tomada de decisão. Avaliação tem papel considerável para desempenhar, mas não pode substituir a política.

Sob um aspecto, contudo, é preciso modular a conclusão de Lundqvist na citação acima. Em vez de recorrer diretamente ao “poder para fazer passar a sua vontade” em casos de desunião, recomenda-se que detentores do poder salientem os traços orientados à conversação, as características discursivas e deliberativas em uma forma democrática de governo. “Poder para fazer passar a sua vontade” é o último refúgio caso diferenças permaneçam depois de extensas deliberações e negociações entre as partes em discordância.

Estratégia de Organização- Há outra solução para o problema do critério que é raramente tratada na teoria de avaliação. Essa solução sugere que o interesse deve focar, não nos resultados diretamente, mas nas instituições por tomarem decisões a partir dos resultados. Essa solução institucional para o problema do critério sugere que a ênfase principal deve ser em erigir instituições que conduzam as avaliações. Questões relativas a quem deve participar de uma avaliação e outras regras que tratam o empreendimento da avaliação como central. O esquema do *stakeholder* que a Suécia institucionalizou no modelo SOU anterior seria o caso em questão. (Veja VEDUNG, 1997, p. 76-83).

A solução institucional implica que o interesse se transfere da aferição cuidadosa de resultados secundários (*outcomes*) e resultados finais para uma aferição meticulosa de arranjos organizacionais e sistemas de regras e normas para a tomada de decisão. Isso marca um retorno irônico da gestão orientada ao processo e modelos institucionais de avaliação, o que a prática contemporânea da avaliação tinha reagido contra: a ênfase em gestão orientada aos resultados e seu estudo direto do resultado primário (*output*) e secundário (*outcome*) da intervenção.

Autoavaliação é outra forma. Assim, o próprio funcionário avalia sua própria performance ou funcionários de uma organização avaliam juntos o desempenho de tal organização. É comum a autoavaliação ser combinada com uma crítica externa por pares e leituras e releituras de rascunhos escritos e comentários uns dos outros, assim como recorrentes conversas entre avaliadores e avaliados.

Modelos de colegiado são constantemente utilizados em avaliação de pesquisa e ensino superior. Eles são praticamente baseados no diálogo, discussão e deliberação. O procedimento usualmente começa com uma autoavaliação. Os profissionais que serão avaliados, fazem uma aferição de seu próprio desempenho, de

seus projetos de pesquisa, programas de educação ou departamento universitário. Eles discutem pontos fortes, fracos, oportunidades e ameaças (usando o chamado esquema *SWOT*). Assim, renomados cientistas externos e educadores da área em particular são designados para aferir a qualidade e relevância do trabalho do avaliado. Esses pares baseiam sua aferição preliminar nas avaliações escritas pelos avaliados e na informação recolhida durante visitas aos departamentos pertinentes através de apresentações orais dos avaliados, conversas com estes e observações. Depois pedem aos avaliados que forneçam comentários escritos sobre o relatório dos avaliadores, e é esperado que os avaliadores leiam e ponderem sobre os comentários, quando eles finalizarem seus relatórios; no final, os avaliadores e avaliados leem os rascunhos uns dos outros, falam um com o outro repetidamente e solicitam a opinião um do outro. Eles se envolvem em algum tipo de processo dialógico argumentativo.

ABORDAGEM DOS OITO PROBLEMAS EM AVALIAÇÃO DE POLÍTICAS PÚBLICAS

Avaliação é uma técnica de gestão de organizações públicas em que a seleção e análise detalhada dos problemas a serem examinados e de questões a serem perguntadas, dados que serão colhidos, análises que serão feitas e outras operações relacionadas à pesquisa têm um papel essencial. Deste ponto de vista, a avaliação baseada na pesquisa se difere da ciência social básica, não em relação ao modelo de pesquisa e ao método de coleta de dados empregado, mas sim aos problemas que serão abordados. Problemas, não modelos ou métodos, dão identidade à avaliação.

Há oito problemas primários de avaliação; eles podem ser colocados como oito questões, ou pontos. A Abordagem dos Oito Problemas em Avaliação de Políticas Públicas ou OPA estão ilustrados na figura 7.1.

Figura 7.1

Abordagem dos Oito Problemas em Avaliação de Políticas Públicas (OPA)

1. Problema do propósito básico *Propósito da questão.* Qual é o problema fundamental da questão que será tratada na avaliação?

2. Problema do propósito superior Independente da questão, é a avaliação lançada por razões de 2.1) prestação de contas, 2.2) desenvolvimento, ou 2.3) pesquisa básica ou será 2.4) pensado como a Vila Potemkin¹?

Fig. 7.1 Continua na página seguinte

¹ Avaliações com cenas Potemkin - Do sueco "Potemkinkulisser" (Fachadas Potemkin, aldeias, vilas). Metáfora ligada à visita da Imperatriz Catarina II à Criméia em 1787, onde o Governador Geral da Ucrânia, Príncipe Potemkin, mandou construir fachadas de casas inexistentes ao longo da rota para dar a falsa impressão de boas condições de habitação. Durante a visita, o Governador Potemkin, ex-amante preferido da Imperatriz, recebeu o título de Príncipe de Tauris.

3. Problema organizacional (avaliador) Quem deve aplicar a avaliação e como deve ser organizada?

4. Problema do objeto de avaliação Como o avaliando, *i.e.*, o objeto de avaliação, a intervenção pública, pode ser delimitado e caracterizado?

5. Problema de implementação

5.1 *Descrição*. Se considerado um meio para algum outro fim (e não uma entidade autossuficiente) como a execução da intervenção pública se apresenta desde a sua instigação formal até os *outputs* finais?

5.2 *Causa*. Quais são os facilitadores e obstáculos ativos na implementação e seu contexto, e como eles influenciam a implementação (análise *upstream*², corolário ao 5.1)?

6. Problema do outcome

6.1 *Outcome bruto*. O que são *outcomes* brutos – imediatos, intermediários e definitivos, dentro e fora da área analisada? *Outcomes* brutos = todos os *outcomes* incluindo fenômenos não causados pela intervenção.

6.2 *Outcome líquido*. Quais são os *outcomes* líquidos da intervenção? *Outcomes* líquidos = aquelas partes dos *outcomes* brutos, dentro e fora da área analisada, que são de certa forma, direta ou indiretamente afetados por pelo menos algum ingrediente da intervenção (análise de efeitos, análise de impactos, corolário a 6.1)

7. Problema do valor

7.1 *Considerações de princípios*. Valor descritivo ou prescritivo?

7.2 *Considerações de princípios*. Antes-do-fato, durante-o-fato ou após-o-fato? Quando o critério de avaliação deve ser determinado?

7.3 *Considerações de princípios*. Por quais critérios valorativos específicos e seus padrões de desempenho nestes critérios valorativos deve o valor da avaliação ser aferido?

7.4. *Aplicação prática*. Quais são os reais valores do avaliando em cada critério e seu padrão?

7.5. *Aplicação prática*. Qual é o composto relevante ou valor real global do avaliando?

8. Problema de utilização (utilidade) Para quem (por quem) e em que modos a avaliação pode ser útil (e utilizada)?

² Análise *upstream* seria uma análise que se inicia no fim da cadeia de implementação e termina no começo, analisando os fatores que facilitam e dificultam.

Optar por oito problemas engendra uma quebra com o mágico número sete. Na cultura ocidental, várias coisas são indicadas pelo dígito sete: as maravilhas do mundo antigo, os pecados capitais, as filhas de Atlas em Plêiades, as eras do homem, os níveis do inferno, os dias da semana. Além disso, psicólogos influentes tem argumentado que o homem é incapaz de processar mais do que sete “pedaços” de informação ao mesmo tempo. No entanto, há conforto no fato de que o título completo do famoso trabalho de George M. Miller sobre nossos limites de processamento de informação é *The Magical Number Seven, Plus or Minus Two* (1956); assim, sua teoria acomoda oito “pedaços” da mesma forma³.

Os oito problemas são arrumados numa fraca ordem lógica. A implicação não é que eles tenham que atender a essa sequência em particular. O avaliador não precisa começar pelo *propósito básico* e então seguir para o *propósito superior*, e assim continuar com os itens 3 a 8. O planejamento de todo o empreendimento, de 1 a 8, pode muito bem começar com o nº 8 - *utilização*, seguir para o nº 03 - *problema organizacional*, para o nº 1 - *o propósito da questão*, e para o nº 7 - problema do valor, então coletar alguns dados no nº 4 e nº 5, e então voltar para o nº 1 - *problema básico* para fazer algumas reformulações. O processo se desdobra em todas as direções, zigue-zague, com uma interação sobre a outra (CRESSWELL, 2013, p. 129 et seq.). Não será feito um esforço aqui para fazer uma representação ideal disso. Em vez disso, apenas comentários sobre os oito problemas em ordem numérica.

Nenhuma questão na Abordagem dos Oito Problemas em Avaliação de Políticas Públicas (OPA) é mais significativa do que a número 1, *problema básico*. Que tópico substantivo será investigado? Como pode ser reformulado em questões investigativas específicas? A primeira parte do capítulo 8 cobrirá isso.

Se a avaliação é comissionada, o avaliador é aconselhado a considerar o ponto nº 2, *problema do propósito superior*. Com isso, quer-se ter algo mais abrangente em mente do que é coberto pelo propósito básico sob o nº 1. O patrocinador quer algumas perguntas esclarecidas sobre a questão-tema, isto é claro, mas

³ Por esta informação Evert Vedung agradece ao amigo e colega Osvaldo Feinstein da Universidade Complutense. (email t EV, 20131118)

por que ele quer isto? Na segunda parte do capítulo 7, quatro propósitos superiores serão discutidos: 2.1) prestação de contas, 2.2) aperfeiçoamento, 2.3) conhecimento básico e 2.4) vilas Potemkin.

Como deve ser *organizado* o trabalho de avaliação? Quem deve arcar com a responsabilidade final da avaliação? Deve ser interna? Ou externa? Ou deve ser uma combinação, interna e externa? Esse 3º problema na Abordagem dos Oito Problemas em Avaliação (OPA) é tratado no capítulo 8.

Antes, o avaliador deve se familiarizar com o *problema do objeto de avaliação*. Como a intervenção deve ser delimitada e descrita? Isso é coberto no capítulo 9 e no capítulo 10.

Se a intervenção pública é tratada como um meio para alcançar algo mais, o que sua execução parece de sua adoção formal até seus outputs finais? Assim o problema 5.1 *descrição da implementação* está formulado na lista da OPA. Associado a isto está o problema 5.2, *causa da implementação*. O avaliador pergunta: o que facilita ou causa impedimentos ou obstáculos à fiel implementação?

O *problema do outcome* é dividido em dois. A primeira parte, número 6.1, refere-se ao mapeamento do *outcome bruto*, *i.e.*, tudo que ocorre nos estágios imediato, intermediário e definitivo. Aqui, a tarefa do avaliador é descrever o *outcome*, não para vindicar até que ponto é causado pela intervenção e sua implementação. Questões feitas por avaliadores nesse estágio incluem: quais pessoas são beneficiários dos outputs? Como eles agem e reagem? O que acontece nos vários estágios do *outcome* além da ação dos beneficiários recipientes? O mapeamento pode incluir também *outcomes* colaterais suspeitos. A perspectiva de tempo é crucial. Dados com uma série temporal estendida nos indicadores de *outcomes* são louváveis.

Desde que a questão seja mapeamento do *outcome* bruto, pode-se chamar de monitoramento. Porém, monitoramento pode produzir material que pode ser usado para investigar o problema número 6.2, *outcome líquido*. Teria a intervenção e sua implementação, ou algum ingrediente do mesmo em algum aspecto e em certa medida, afetado uma parte dos *outcomes* brutos mapeados (=efeitos líquidos)? *Outcome líquido* ou impacto não é um assunto fácil de se explicar. A intervenção através de alguns mecanismos produziu o *outcome* alvo? Facilitou ou combateu o desenvolvimento?

Precipitou o *outcome* alvo, ou o retardou, ou não gerou nenhum efeito perceptível? Foram produzidos *outcomes* fora da área analisada, *i.e.*, efeitos colaterais? Para responder a essas perguntas difíceis, a comunidade científica desenvolveu uma família de modelos, a ser discutido nos capítulos 12-14.

O item 7 é sobre *valor ou valoração*, uma questão vital na avaliação. O inútil deve ser separado do estimável e digno, o trigo deve ser peneirado do joio. O problema da valoração é dividido em três considerações de princípios e duas aplicações práticas. Valoração já foi coberta no final do capítulo 5.

Finalmente, o avaliador deve confrontar o problema da *utilidade ou utilização da avaliação*. Para quem a avaliação será útil? Quem é o público-alvo? Pode ser útil e usado em outras situações? Deve a avaliação ser organizada ou delineada de forma a melhorar a utilidade e uso?

A lógica de avaliação é diferente daquela da pesquisa básica. Pesquisadores básicos trabalham como seus próprios empresários, sem utilidade prática em vista. Os avaliadores, ao contrário, muitas vezes trabalham para os contratantes e os informam, eles por sua vez afirmam publicamente que querem resultados da avaliação para uso em um contexto prático iminente. Isso faz com que o tempo, o *feedback*, a utilidade e a utilização sejam cruciais para a avaliação, mas nem tanto para a pesquisa básica; conseqüentemente, o uso é uma questão muito debatida em avaliação, mas raramente abordado na ciência social pura.

PROPÓSITO BÁSICO, PROPÓSITOS SUPERIORES

O propósito da avaliação não é provar, mas melhorar.

*Egon Guba – frase extraída de Daniel J. Stufflebeam, –
The 21st Century CIPP-model (2004, p. 262)*

O propósito mais importante da avaliação
não é provar, mas melhorar.

Daniel J. Stufflebeam - The 21st Century CIPP-model (2004, p. 262)

“Problemas são as fontes da pesquisa científica”, assim diz o ditado. O ponto de partida lógico para todo questionamento científico é um problema. Métodos, teorias e dados não são de primeira importância; problemas são. Isso também se aplica à avaliação. O propósito avaliativo básico, de tópico definido, deve ser fraseado e aguçado, preferivelmente na forma de pergunta(s). Esse é o *problema do propósito básico*, o primeiro na nossa Abordagem dos Oito Problemas em Avaliação de Políticas Públicas (OPA). O segundo problema – para quais *propósitos superiores* a avaliação é lançada? – também será trabalhado neste capítulo.

8.1 Conceitos: O Problema do Propósito Básico e dos Propósitos Superiores

Propósito Básico

O que exatamente deve ser avaliado, qual o *propósito básico* e que perguntas devem ser perguntadas? Como a intervenção a ser escrutinada deveria ser circunscrita e representada? Talvez o contratante da avaliação tenha formulado suas intenções em *termos de referência* (TDR). Esse é o procedimento operacional padrão das avaliações de cooperação para o desenvolvimento da Comissão Europeia e da Agência Sueca de Cooperação Internacional para o Desenvolvimento (SIDA). Como segundo passo, a União Europeia (UE)

requer um *Inception Report*¹. Nesse documento escrito, espera-se que os avaliadores prospectivos indiquem suas interpretações das questões avaliativas contidas no Termos de Referência (TDR) e seus delineamentos do assunto, bem como suas escolhas do projeto e dos métodos de coleta de dados, como também um esboço do plano de execução da avaliação como um todo. Esse relatório preliminar forma a base para o diálogo entre o comprador e o fornecedor prospectivo. Caso tais procedimentos escritos não se apliquem, provavelmente haverá um ocasional uma conversa detalhada entre contratante e avaliador. Essa procura pelo problema de propósito da questão básica (fundamental) a ser sondada pode ser reforçada com um estudo de viabilidade, seguido de mais conversas antes de um acordo ser fechado. Na realidade, a lapidação do propósito da questão básica continuará durante todo o processo de avaliação. Através da confrontação de percepções com dados empíricos, o avaliador faz repetidas descobertas, trazendo a necessidade de novas questões e o refinamento do problema básico em subproblemas. Esse ajuste do propósito básico pode se tornar uma operação contínua.

Questões chave a serem levantadas logo de partida são: a avaliação deve se focar em *outcomes* ou em processos? Os custos também devem ser levados em conta?

Considere um acordo em que o avaliador deve providenciar informações a respeito de uma recente política de segurança no trânsito em estradas públicas. Fatalidades relativas ao trânsito têm diminuído continuamente nos últimos três anos. Isso se deve à política de segurança nas estradas, vigente há quatro anos, ou a algo de ocorrência simultânea? Sendo o caso a segunda opção, a quê? Se devido à nova política, como os efeitos provocados deveriam ser julgados? Que efeitos colaterais isso tem sobre a mobilidade dos trabalhadores que se locomovem de carro e como isso afeta, por sua vez, a propensão das pessoas a viver em áreas rurais fora das cidades onde trabalham? Que elementos da política foram eficazes (e ineficazes) dentro e fora da área alvo? Ademais, as partes concordaram em incluir processos de implementação entre as atribuições presentes no contrato.

¹ Relatório de avaliação preliminar entregue à UE dentro de 9 meses do início do projeto. *Inception Report Template*

Em suma, a formulação preliminar da questão se dá da seguinte forma: 1) a avaliação deve estimar os impactos resultantes, efeitos colaterais incluídos, da nova política de segurança no trânsito, 2) componentes efetivos da política devem ser discernidos e estimados e 3) a implementação deve ser incluída entre os mecanismos potencialmente causais².

Propósitos Superiores

Para quais *propósitos superiores* a avaliação deve ser lançada? Assim é formulada a segunda pergunta na lista dos OPA. A avaliação é executada para: 1) *accountability*, 2) melhoria, 3) progresso de conhecimento básico ou 4) propósitos estratégicos³. Começaremos pelos três propósitos substantivos e terminaremos pelos estratégicos⁴.

8.2 Accountability: O Agente Presta Conta, O Principal Afere

A ideia central da avaliação de *accountability* é descobrir se o agente exercitou o poder delegado a ele e cumpriu seus deveres de forma apropriada, de modo que o principal possa julgar seu trabalho, tomar medidas corretivas ou conceder quitação. A *accountability* (controle pós-fato) envolve duas partes, o principal e seu agente, e duas funções, a *accountability do agente* num sentido e a *aferição do principal* no outro. O principal não pode cuidar de tudo sozinho, recorrendo a delegar atribuições a um agente; este concorda então em assumir esses deveres.

² Pontos interessantes de análise de problema em Bryman (2008, p. 65 et seq).

³ No nível mais alto de abstração, toda avaliação trata de *aprendizado*. A avaliação reflete em algumas atividades e eventos de modo a destilar lições a serem aplicadas no pensamento e ação futuros. Esse tema não mais será abordado aqui.

⁴ A maioria dos especialistas trabalha com três propósitos, deixando de lado o quarto propósito, o estratégico (Vilas de Potemkin). Em verdade, Joe Hudson, John Mayne e Ray Thomlison (1992, p. 5) têm quatro propósitos: ganho de conhecimento, melhor execução de programas, redirecionamento de programas e fornecimento de dados para *accountability*. Entretanto, melhor execução de programas e redirecionamento de programas podem ser resumidos em melhorias, fazendo com que se possa reduzir seu esquema a três propósitos, dentre os quais o estratégico continua ausente.

Espera-se, após algum tempo, que o agente (prestador de contas) preste conta de suas atividades ao principal (responsável, *accountee*). Com base nisso, o principal pode controlar se o agente seguiu o acordado, dar-lhe crédito e conceder quitação ou criticá-lo e tomar ações corretivas. Entretanto, a prestação de contas também pode ser executada por um corpo externo, comissionado pelo principal, ou pelo próprio principal. Essencialmente, a avaliação para *accountability* pretende servir aos interesses de um corpo externo supervisor.

A avaliação de *accountability* pode se focar em diversos aspectos do objeto de avaliação (avaliando):

- | | |
|-----------------------------------|--|
| 1 Aferição de Impacto: | A intervenção vem produzindo o <i>outcome</i> esperado? |
| 2 Aferição de Cobertura: | As pessoas contempladas são o público alvo designado? Há beneficiários que não deveriam ser contemplados? |
| 3 Aferição de Cumprimento: | A emissão de <i>outputs</i> vem ocorrendo em quantidades satisfatórias? O <i>output</i> emitido corresponde ao que deve cumprir a intervenção? |
| 4 Aferição Fiscal: | Os fundos estão sendo utilizados de maneira apropriada? Os gastos vêm sendo corretamente documentados? Os fundos vêm sendo gastos dentro dos limites do orçamento? |
| 5 Aferição de Eficiência: | Como se comparam os impactos em relação aos custos de intervenção? |
| 6 Aferição Legal: | As leis relevantes vêm sendo cumpridas, incluindo aquelas que dizem respeito à saúde e à segurança ocupacionais, representação da comunidade em conselhos decisórios, provisão imparcial de serviços, consentimento bem informado e privacidade nos registros individuais?
(ROSSI & FREEMAN, 1989, p.157 et seq.) |

O inventário de seis tipos exhibe tipos de informações a serem reunidos na avaliação de *accountability*⁵.

⁵ A avaliação de *accountability* (*responsabilização*) é similar, mas não idêntica, à avaliação cumulativa na terminologia de Michael Scriven (1991, p.340). A similaridade é que ambas são conduzidas em benefício de uma audiência externa. A diferença parece ser que essa pode se tratar de qualquer audiência externa no segundo caso, enquanto no primeiro, ela deve se tratar de uma audiência superior. A avaliação de *accountability* está associada a noções de supervisão, hierarquia administrativa e direito de saber, na interpretação deste livro.

8.3 Quatro Perspectivas de *Accountability*

"A *accountability* é a ligação entre a burocracia e a democracia [representativa]. A democracia [representativa] moderna depende da *accountability* das burocracias para tocar a política declarada e ademais administrar as estruturas vigentes de oportunidade e regulação determinadas pelo governo". Essas palavras (com adendos), proferidas por Lipsky (1980), contém uma verdade profunda, que deveria ecoar mais na literatura sobre avaliação do setor público.

Na avaliação de *accountability* (*responsabilização*), a *perspectiva do funcionário político* (*cargos de confiança*) geralmente é tida como fundamental. Os políticos requerem a avaliação para responsabilizar a administração por suas ações. Contido nesse dever está o que Bo Rothstein (1997, p.48 et seq.) chamou de "dilema fundamental da administração pública".

Corpos governativos eleitos, como parlamentos e câmaras municipais, só podem tomar uma pequena parcela de todas as decisões que dizem respeito ao bem-estar dos cidadãos, feitas em sociedades altamente desenvolvidas. As razões, segundo Rothstein, são duas: a falta de competência e a limitação do tempo disponível.

A maioria esmagadora das resoluções no âmbito do setor público é concernente a instâncias individuais ou a situações específicas. Temos, por exemplo, a colocação de pacientes em filas de espera por cirurgias, as decisões de professores a respeito das notas de seus alunos ou a alocação de empréstimos de habitação do governo para empreiteiras de construção civil individuais. Frequentemente, essas decisões devem ser tomadas constantemente, como pelo médico em sua clínica, pelo professor em suas aulas ou pelo profissional de linha de frente que atende sua clientela de asilos municipais. As resoluções requerem conhecimento muito mais especializado das situações específicas do que os membros democraticamente eleitos de uma assembleia podem razoavelmente possuir. *Falta-lhes competência*. A necessidade de adaptação situacional no que diz respeito à qualidade da informação do destinatário é tão exigente que os funcionários eleitos não têm competência para exercer qualquer influência detalhada sobre esse tipo de decisão.

Da mesma forma, as *limitações de tempo* tornam impossível para os políticos eleitos aprender as peculiaridades necessárias a cada item do setor público. Geralmente, eles só tomam uma decisão fundamental numa mesma área substantiva a cada dois, três ou quatro anos.

Entretanto, na maior parte dos domínios de política funcional, decisões devem ser tomadas várias vezes a cada mês, semana e, ocasionalmente, até dia, continuamente. As resoluções parlamentares nesses casos, portanto, devem ser gerais e apenas conter diretrizes. Isso, por sua vez, significa que muitas decisões extremamente importantes para os cidadãos ou clientes devem ser tomadas incessantemente por corpos compostos de membros não diretamente eleitos pelo povo, mas que tem poderes delegados a eles por instituições parlamentares ou por autoridades que por sua vez tem poderes delegados a elas por instituições parlamentares. Instituições democraticamente eleitas não tem influência formal sobre as decisões tomadas acerca desses assuntos.

A falta de competência e limitações de tempo forçam as assembleias parlamentares a delegar seu direito de decisão em áreas fulcrais aos governos, e esses por sua vez a outras instituições, preferivelmente à administração pública. Entretanto, essas decisões são políticas, uma vez que elas influenciam diretamente a distribuição de valores na sociedade, isso é quem recebe o quê, quando e como. Essa delegação de poder mudou o foco da administração da execução de políticas para a tomada de decisões de cunho político. O servidor ainda é um servidor, mas seu poder de decisão aumentou, uma vez que seu mestre assumiu mais e maiores responsabilidades.

Nessa *perspectiva política*, a avaliação se torna uma importante ferramenta para que o principal possa confirmar o cumprimento das atribuições designadas ao agente administrativo e, se necessário, tomar as devidas providências.

Na *perspectiva da gestão de agência*, a liderança da agência é encarregada da necessidade legítima de ficar de olho em seus subordinados. Bem como os políticos, mesmo esses principais devem delegar seu poder a subordinados de menor nível dentro da mesma hierarquia, a profissionais autônomos, como juízes, médicos, assistentes sociais e professores, a partes independentes não governamentais como organizações de interesse, ou a consultores e clientes externos. A avaliação se torna uma ferramenta para os chefes de agência se manter informados sobre o manejo de tarefas delegadas e o desempenho de seus subordinados.

Esses dois aspectos da *accountability* geralmente são vistos como os fundamentais. Entretanto, existem ao menos mais dois ângulos que devem ser trazidos à tona.

Um deles é a *perspectiva do cidadão*. Embora os cidadãos sejam o ponto de partida da corrente parlamentar de governança – em níveis nacional, regional, municipal e europeu – e todos os outros atores na cadeia sejam seus agentes, a perspectiva do cidadão raramente é ouvida no discurso sobre avaliação. Mas os cidadãos precisam de fato da avaliação, para que possam aferir o desempenho dos funcionários eleitos e de seus agentes em diversos níveis e responsabilizá-los nas próximas eleições. “Todo o poder emana do povo e o parlamento representa o povo sueco”. Essas são palavras da Constituição Sueca de 1974⁶. Segundo a teoria da democracia representativa, os políticos são os representantes do povo. Eles são eleitos pelo povo para fazer valer a sua vontade (o seu desejo), dentro de certos limites discricionários. O povo é o principal e os políticos eleitos são seus agentes. Os funcionários políticos, a teoria continua, tomarão então decisões em acordo com a constituição. Informação sobre a implementação, *outputs* e *outcomes* dessas decisões deve ser crucial quando, nas próximas eleições, os cidadãos forem avaliar as políticas passadas para exercer seu poder de eleger novos representantes.

Também existe a *perspectiva do cliente* na avaliação. Clientes constituem uma categoria mais limitada do que cidadãos, mas seu interesse em que os programas produzam algo mais do que apenas palavras, que casos iguais sejam tratados equitativamente e que o serviço seja satisfatório também é legítimo. Avaliações desses aspectos se tornarão muito importantes para eles.

8.4 Armadilhas Sociais na Avaliação de Accountability

A *accountability* (responsabilização) é ameaçadora para um agente. Ele teme críticas adversas que possam manchar sua reputação ou, ainda pior, tirar seu emprego. Mas os principais, de todas as pessoas, não deveriam estar curiosos sobre como o seu subordinado executa suas tarefas?

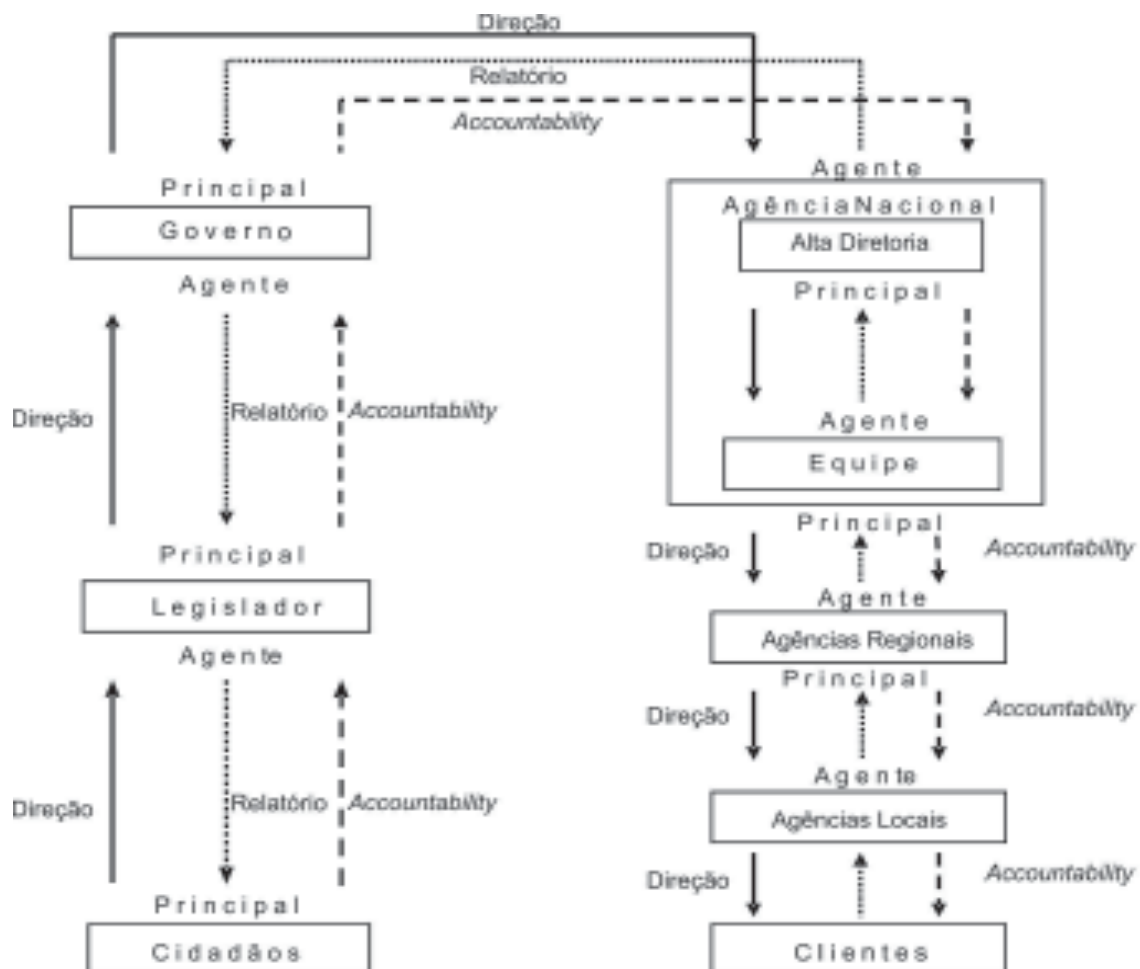
Esse argumento não leva em conta um dilema enfrentado pelo principal: a *tiranía do duplo papel*. Na vida pública, principais são simultaneamente superiores e subordinados, amos e servos.

⁶ Na Constituição da República Federativa do Brasil de 1988, 35ª Edição (2012) lê-se: “Artigo 1º, parágrafo único. Todo o poder emana do povo, que o exerce por meio de representantes eleitos ou diretamente, nos termos desta Constituição.”.

A burocracia nacional é o principal (o “chefe”) das agências regionais em suas jurisdições, mas é também o agente do seu principal, o governo. E o governo, por sua vez, é o principal da burocracia nacional, mas o agente da legislação. A mesma instituição é, ao mesmo tempo, Dr. Jekyll e Mr. Hyde. Por conta de seu duplo papel, o principal pode se deparar com situações em que a avaliação de *accountability* de seus funcionários é usada como inspeção dele mesmo (principal) por seus superiores.

Avaliações de *accountability* que tem por finalidade servir às necessidades de informação do principal podem acabar se virando contra ele mesmo. Esse medo de se tornar não intencionalmente responsabilizável provoca a uma forte disposição de evitar avaliações como um todo. Do mesmo modo, caso uma avaliação revele resultados muito negativos, a confecção de um relatório será ativamente reprimida. Se o relatório vier a público ainda assim, sua disseminação será freada. Se o relatório se dissemina, tenta-se ocultá-lo, distorcer ou esconder as informações negativas e engrandecer as positivas. A avaliação se torna uma luta nua e crua pelo poder. O argumento está ilustrado na figura 8.1.

Figura 8.1
Duplo Papel na Cadeia de Representação Pública



Nota. O principal dirige seu agente e o agente é obrigado a reportar de volta para o principal, para dar a este uma base de informação sobre a qual responsabilizar seu agente.
Direção = —————> Relatório => Accountability = - - - ->

Para alguns indivíduos, a tirania do duplo papel é ainda mais saliente. Legisladores eleitos, por exemplo, às vezes são apontados pelo governo para compor os conselhos das comissões reguladoras, o que significa que em uma capacidade eles devem atuar enquanto agentes do governo, ao mesmo tempo em que personificam o principal gestor público.

A segunda razão pela qual a avaliação de *accountability* é evitada pode ser expressa como o *poder da accountability (responsabilização) de corroer a confiança*. A avaliação de *accountability* pode se tornar contraproducente. Assume-se que a *accountability* produz maiores níveis de eficiência e conformidade no atendimento das demandas do principal, mas o resultado pode ser inesperadamente o contrário. Uma ênfase forte demais na *accountability* pode levar o agente a adotar uma abordagem covarde em relação às iniciativas. Conhecendo cada capricho e hábito de seu mestre, Passepartout, aquele servo perfeito de “Volta ao Mundo em 80 Dias”, de Júlio Verne, antecipava os desejos de Phileas Fogg ao ponto deste não precisar dar uma única ordem. Sua relação funcionava como se houvesse – para pegar emprestada a frase memorável de Arthur Okun – “um aperto de mão invisível” entre os dois, levando o mestre a oferecer “o pagamento justo por um dia de trabalho” e o servo a “dar um dia de trabalho honesto”. Um destaque muito grande para a *accountability (responsabilização / responsabilidade)* pode transformar essa inclinação para a antecipação benigna em subserviência covarde e condescendência.

Tanto a tirania do duplo papel quanto o poder da *accountability* de corroer a confiança, cada um à sua maneira, contribuem por um interesse de baixo a moderado na *accountability (responsabilização / em chamar à responsabilidade)* na vida pública.

A ansiedade que a *accountability* causa no principal, entretanto, não é nada se comparada à apreensão sentida pelo agente. Eles temem a crítica por parte de seus superiores, que podem manchar reputações e ameaçar empregos. Ele, o agente, rodeia, se recusa a cooperar, divulgando apenas material favorável à sua causa. Ele desacredita as descobertas do avaliador como ataques interesseiros, paroquiais e ingratos à sua probidade e eficiência. Avaliadores e contratantes de avaliação são vistos como antagonistas, senão mesmo como inimigos. Isso também explica em parte por que a função da avaliação é tão fraca na vida pública.

A *teoria econômica do principal-agente* sugere outra explicação para o desconforto do agente com a avaliação de *accountability*. Ele age de forma desonesta em relação a seu principal devido à *assimetria de informação* em seu favor.

O agente costuma agir segundo os seus próprios interesses. Ele cumpre o suficiente das suas obrigações para passar pelo crivo de seu principal, mas nada mais. Ao invés disso, ele se dedica a assuntos privados, viaja para conferências, manda cartas particulares com material do escritório, faz ligações particulares com seu telefone de serviço, copia documentos de família nas copiadoras do escritório, envia mensagens de e-mail particulares pela rede de seu empregador em horário de trabalho. A razão pela qual o agente pode fazer isso é a sua enorme superioridade de informação. O funcionário público (agente), que, entra dia e sai dia, lida com os clientes, sabe infinitamente mais sobre o trabalho que está sendo de fato realizado do que o chefe da repartição pública (principal), que nunca encontra com cliente algum. O professor (agente) que passa horas todos os dias instruindo os estudantes, está marcadamente melhor informado sobre as atividades, num nível fundamental, do que o diretor da escola (principal), que raramente os vê numa sala de aula. Essa assimetria de informação vira o jogo em favor do comportamento egoísta do agente.

Esse comportamento egoísta é *estratégico*, uma vez que as pessoas agem baseadas no que elas *acreditam* que os outros vão fazer. Nós nos deparamos com ações baseadas na *antecipação* do futuro, ou seja, as pessoas adiantam as ações das outras, bem como suas consequências, e agem com base nessa previsão. A ação é determinada por *crenças* prescientes no futuro, não por valores. Em sua *accountability*, o agente maquia a realidade porque acredita que, do contrário, seu chefe irá repreender suas ações, ou ao menos culpá-lo de algo. Ele acredita ainda que seus colegas nas demais unidades, que reportam ao mesmo chefe, farão o mesmo. Ele não confia nem em seus colegas nem em seu chefe. Seu comportamento desleal é facilitado pela assimetria de informação. Isso torna a base de informação produzida pela avaliação de *accountability* uma ferramenta de gestão muito mais frágil do que ela seria no melhor dos mundos (POLLITT, 2003, p.88, p.142 et seq.; DOUMA; SCHREUDER, 1998; DUNLEAVY, 1991).

8.5 Melhoria, Desenvolvimento

Na perspectiva da *melhoria* (desenvolvimento, promoção), o estudo é genuinamente dedicado à coleção de fatos sólidos e imparciais sobre os quais possa ser construída a verdade da intervenção, de forma a ser capaz de proporcionar uma base de informação para o melhoramento. Nenhuma intenção duvidosa espreita no pano de fundo.

O propósito da melhoria é formativo ou aditivo. Na melhoria formativa, o avaliador não põe em questão as metas da atividade, mas se concentra nos meios, métodos, estratégias e ações para alcançar essas metas. A avaliação de melhoria formativa é um processo de descoberta de falhas no quadro estrutural existente com o propósito de melhoramento incremental da implementação de intervenções, a fim de melhor alcançar dados objetivos. A intervenção aspira um *aprendizado de ciclo único*, ou seja, aprendizado dentro dos contornos de uma intervenção já adotada. O propósito não é desafiar atributos básicos como objetivos e problemas subjacentes, mas guiar o refinamento e o alinhamento da implementação da intervenção. A pergunta fundamental é: “Como a implementação pode ser melhorada?”

A *avaliação de melhoria aditiva*, por outro lado, envolve a *reconsideração da intervenção*. O propósito é o *aprendizado de ciclo duplo*, no sentido em que a decisão fundamental de intervenção está sujeita à possível modificação. A direção fundamental da intervenção em si está em jogo. A avaliação aditiva questiona não só os meios, métodos e estratégias, bem como os objetivos e talvez até mesmo o problema que subjaz a intervenção. A ideia é tirar conclusões de intervenções (quase) finalizadas a fim de acrescentar algo novo. Ambos o aprendizado de ciclo único e o de ciclo duplo estão incluídos no propósito superior de melhoria compreensiva. Pondo de outro modo, avaliações podem ser dirigidas por *desenvolvimento* e *promoção formativos* bem como pelos *aditivos*.

A avaliação para melhoria é frequentemente conduzida como *autoavaliação*. O operador *street-level* (de nível de rua) ou a gerência de agência pode autocriticamente aferir seu próprio desempenho. Ainda assim, a melhoria também pode se dar por *avaliação externa*. Idealmente, a tirania do duplo papel pode não se manifestar aqui,

uma vez que o propósito expresso é a melhoria indistinta onde ninguém está jogando nenhum jogo estratégico contra outro operador ou gerente.

A *accountability* e ambas as melhorias formativa e aditiva são alvos dignos da avaliação.

8.6 Avaliação para Conhecimento Básico

O terceiro propósito elevado é o *conhecimento básico*. A avaliação agora é vista como pesquisa fundamental para aumentar a compreensão geral e ultracasuística de instituições, organizações, processos, atividades, eventos, estratégias de governança, instrumentos de política ou a forma como funcionam as agências (FRANKE-WIKBERG; LUNDGREN, 1980, p.129 et seq., CHEN, 1990, e FITZ-GIBBON; MORRIS, 1975). Ela aumenta o corpo de conhecimento amalgamado de algum campo acadêmico de estudo. Mesmo aqui, há um propósito de melhoria, de certa forma, embora engendrando um melhoramento de longo-prazo, mais amplo e distante (RUTMAN; MOWBRAY, 1983, p.27 et seq.).

A perspectiva do conhecimento básico é visível na *meta-avaliação* e nas *revisões sistemáticas*, interpretadas como atividades de sumarização de todo um leque de avaliações individuais de uma dada área funcional. O ponto em tais inventários sumários deve ser extrair as ideias gerais de uma gama de estudos separados através de sua integração em estruturas conceituais mais abrangentes, de forma a elevar seu valor para tomadores de decisão, bem como para pesquisadores básicos (BERNSTEIN; FREEMAN, 1975; ROSENTHAL, 1984; LIGHT; PILLEMER, 1984; VEDUNG, 1982a).

8.7 Propósitos Estratégicos (“Vilas de Potemkin”)

Conduzida em cenários de ação, a avaliação é usualmente permeada de considerações orientadas a jogos, propósitos *estratégicos*. Pessoas que agem estrategicamente agem com base no que elas *acreditam* que os *outros* farão. A ação estratégica é fundada na *antecipação*, ou seja, as pessoas preveem as ações futuras de outras pessoas e suas prováveis consequências e agem com base no que elas acreditam que isso representará para elas.

No modo estratégico, avaliações podem ser comparadas a “Vilas de Potemkin”, ou simplesmente “Vilas Potemkin”. A metáfora representa dispositivos aplicados pelo agente para camuflar más condições para seu principal. A expressão se refere à famosa visita da imperatriz russa Catarina II à Crimeia, em 1787, quando o Príncipe Potemkin, governante geral da Ucrânia, construiu fachadas de casa artificiais ao longo do trajeto, para dar a ela a falsa impressão de que a província sob seu governo havia prosperado. Grigori Potemkin – o agente –, até então o amante favorito da imperatriz, que havia conquistado e governado a província, tentou usar de sua superioridade de informação sobre a imperatriz – o principal –, que, na distante São Petersburgo, não tinha a possibilidade de se informar sobre os problemas da área. O motivo por trás de sua tentativa de enganar Catarina foi seu interesse pessoal em manter o favoritismo e evitar qualquer reação negativa por parte dela⁷.

A avaliação dos dias de hoje também pode ter Vilas de Potemkin como força motriz. O executivo estabelece avaliações para esconder de seu principal insuficiências e falhas, veicular imagens atraentes de intervenções e, no geral, prover aparências mais deslumbrantes do que a realidade. O principal comissiona avaliações em situações de crise para ganhar tempo e deixar a poeira baixar, mostrar uma fachada de racionalidade e disseminar uma visão embelezada do trabalho de suas agências. Esses motivos para a avaliação são frequentemente ocultados. Se revelados, perderiam seu pretenso valor no jogo de poder político administrativo (ANDERSON; BALL, 1978, p.34 et seq.).

Gurus dos métodos às vezes condenam considerações estratégicas como abuso de avaliação. Edward Suchman, o pioneiro estadunidense, as descarta como pseudoavaliações (1972, p.81). Sua visão pressupõe que poderíamos ter avaliações inteiramente limpas, purgadas de manobras políticas estratégicas. Isso é utópico. É impossível erradicar a luta pelo poder em grandes organizações, que dirá no Estado.

⁷ “Elefante branco” é uma expressão utilizada na África para designar um projeto de estimação do governo, que vai muito bem e ao qual são levados convidados estrangeiros e avaliadores a fim de que tenham uma impressão exageradamente boa dos resultados alcançados por seus programas de auxílio ao desenvolvimento. “Elefante branco” é a contraparte africana para “Vilas de Potemkin” ou “Fachadas ou Aldeias Potemkin”.

Entretanto, Suchman toca em uma fragilidade da avaliação. A existência de propósitos estratégicos ocultos por trás das palavras bonitas sobre melhorias indica os limites da avaliação enquanto produtora de informação para decisão racional no setor público. A avaliação se dá num cenário político. A política é uma atividade imbuída de valores. Um principal que já se decidiu pela retração se utiliza da avaliação para encontrar argumentos plausíveis para essa linha de ação. Ao invés de relatar fatos objetivos nas prestações de contas a seu principal, o agente maquia a base de dados. Considerações estratégicas reduzem o valor da avaliação como emissora de conhecimento imparcial.

AUTOAVALIAÇÃO, AVALIAÇÃO INTERNA OU EXTERNA

Quem deve realizar a avaliação e como ela deve ser organizada? A primeira parte deste terceiro problema da Abordagem em Oito Problemas pode parecer estranha. Em países democráticos, a avaliação pode ser conduzida por parlamentos, governos ou agências. Cidadãos, jornalistas, clientes, bem como organizações de interesse também podem fazê-la. Aqui, neste capítulo, serão tratados somente alguns aspectos. A autoavaliação é uma coisa razoável a se fazer? Ou a avaliação deveria ser realizada por outra pessoa, de dentro ou de fora da organização pertinente?

Essa não é realmente uma situação de escolha, pode-se argumentar. A avaliação deve ser sempre externa, ou pelo menos por pessoas não diretamente envolvidas na formação, adoção, implementação e recepção do objeto da avaliação. A avaliação interna resulta em vanglória, arrogância, autopropaganda e encobrimento apenas. E ninguém vai aprender a partir dessa avaliação interna.

Apesar de conter um pouco de verdade, essa visão é imprecisa. É necessário primeiro esclarecer o que é "autoavaliação". Quando o operador em sua agência de energia, o professor em sua escola, o juiz em seu tribunal ou o inspetor em sua gestão ambiental municipal avaliam cuidadosamente as suas próprias práticas profissionais, com vista a promover a sua própria melhoria, eles estão, certamente, realizando autoavaliações. Ao deixar o ator individual e entrar no nível coletivo, a classificação se torna mais complicada. No entanto, ainda se estará tratando de casos de autoavaliação se todos os operadores de certo tipo de tarefas na agência, todos os professores de um assunto específico em uma escola, todos os juízes em um tribunal local ou todos os inspetores da gestão ambiental municipal juntos avaliarem seus próprios desempenhos. O futuro da autoavaliação, nesse sentido, parece de fato muito brilhante.

Agora, suponha que o objeto da avaliação seja uma organização formal. Se a avaliação é realizada por alguma outra organização, então é externa. Entretanto, e se a avaliação for realizada por uma

unidade de avaliação dentro da mesma organização?¹ Um exemplo pode ilustrar como essa demarcação vai funcionar.

Em 2010, a Agência Sueca para a Avaliação do Desenvolvimento (SADEV), localizada fora da zona da capital, era uma autoridade independente para a avaliação da assistência ao desenvolvimento gerida pela Agência Sueca de Cooperação Internacional para o Desenvolvimento (*The Swedish International Development Cooperation Agency*, SIDA, em inglês), na cidade de Estocolmo, capital do país. Ao mesmo tempo, dentro da SIDA, havia uma unidade chamada Secretaria de Avaliação. Longe de participar na assistência ao desenvolvimento da SIDA, a Secretaria trabalhou inteiramente voltada para sua própria avaliação. Quando a SADEV realizava avaliações das atividades da SIDA, eram evidentemente externas. Externas foram também as avaliações da cooperação da SIDA executadas por empresas de consultoria com fins lucrativos de Londres, Inglaterra, ou de Strängnäs, Suécia. Por outro lado, uma vez que a Secretaria fez uma avaliação da ajuda ao desenvolvimento da SIDA, seria estranho considerar essa avaliação como externa. Se o limite é pautado na organização formal, então devemos considerá-la como interna. É uma autoavaliação? Ao nível da organização como um todo, sim. Ao nível dos funcionários individualmente, não, pela simples razão de que os funcionários da SADEV não estavam avaliando-se, mas o que as outras unidades e outros funcionários dentro da Sida estavam fazendo. É importante ter essa distinção em mente.

9.1 De Avaliação Interna ou Externa?

Após esta digressão semântica, é retomado o quebra-cabeça: existe lugar para a avaliação interna no setor público ou a avaliação externa é a solução evidente. No primeiro caso, uma posição aparentemente forte pode ser defendida para a autoavaliação, uma vez que, idealmente, no melhor dos mundos, cada organização deve passar pelos processos de autoavaliação e de autoaprendizagem (WILDAVSKY, 1985). "Avaliação interna eficaz fornece um suporte indispensável para os gestores", Arnold Love argumenta em seu livro

¹ Como a função de avaliação pode ser institucionalizada no governo não é abordado no presente trabalho. Como ele está organizado em vários países é tratado em Rist (1990) e em Furubo, Rist e Sandahl (2002).

perspicaz *Internal Evaluation* (1991, p. 5, também SONNICHSEN, 2000). O mesmo argumento pode ser usado para grupos-alvo, profissionais, administradores, políticos eleitos e outras partes interessadas em uma organização. Todos podem aprender com a avaliação interna².

Embora isso seja provavelmente verdade, a avaliação interna não pode resolver todas as necessidades de avaliação, por causa do comportamento de autopreservação disseminado na organização. No entanto, as avaliações também nem sempre podem ser externas à agência. Para alguns propósitos, a avaliação interna é indubitavelmente apropriada. Em outros contextos, a avaliação externa é mais adequada. A solução mais sábia é a escolha no que diz respeito aos objetivos mais elevados da aferição. Isso significa que os três propósitos superiores, sendo eles o de prestação de contas, o de melhoria (desenvolvimento) e o de conhecimentos básicos podem exigir diferentes projetos organizacionais.

Além disso, as avaliações interna e externa podem e devem ser combinadas na mesma avaliação. Em uma avaliação escrita de um curso universitário, os alunos são convidados a avaliar o seu próprio desempenho (autoavaliação), bem como o de seus professores (avaliação externa). A avaliação de pesquisa muitas vezes começa com uma autoavaliação da própria equipe científica relacionada com a disciplina, cujos resultados são posteriormente usados como base de informação para seus pares revisores externos. Suas opiniões preliminares são então retroalimentadas de volta para os pesquisadores avaliados elaborarem seus comentários, e será com base neles que os revisores vão elaborar o seu relatório de avaliação consolidado. Em ambos os casos, a autoavaliação confere legitimidade e apoio para o conjunto da atividade de avaliação. Não se trata de escolher entre a avaliação interna ou a externa, mas escolher ambas. A decisão fica a cargo do propósito superior, mas as escolhas estão longe de serem claras.

² Pedersen et al. (2003) é muito interessante sobre a autoavaliação. Veja também em Holmer, Nilsson, Nilsson, Rovio-Johansson e Åberg (1992), esse sistema interessante de autoavaliação e de avaliação externa que foi construído em Gotemburgo, em Áreas Metropolitanas 1998-2006, são discutidos em Hertting e Vedung (2006, p. 148-158 e 141 et seq.), mas especialmente em Haraldsson, Holmer, Ketola e Krantz (2005, p. 46 et seq.); Esse é também um modelo de autoavaliação.

9.2 Prestação de contas

Quando o propósito de nível superior é a prestação de contas a terceiros e a objetividade é importante, um forte argumento pode ser usado em favor da avaliação externa. De preferência, deve ser conduzida por um corpo autônomo, imparcial. Geralmente, as avaliações externas carregam maior credibilidade como iniciativas objetivas do que as avaliações internas.

Avaliadores internos não podem realizar avaliações imparciais de prestação de contas por duas razões. Se o avaliador interno sente que algum componente do programa funciona bem, ele pode sujeitá-lo a um exame minucioso. Ele pode ater-se aos fatos e considerá-los de uma maneira perfeita. A parcialidade reside nos projetos duvidosos ou aparentemente sem sucesso, que não estão sujeitos ao mesmo escrutínio detalhado. O avaliador conhece pessoalmente os responsáveis do componente do programa e sabe que ele é mal gerido. Mas o avaliador hesita em expor isso para conhecimento geral. Ele pode reunir os dados negativos em um relatório, mas de uma maneira otimista e casual, a fim de ter uma referência. Contudo, para evitar o pesado fardo de revelar deficiências, ele não investigará o assunto com muito cuidado.

Além disso, se o projeto custou muito a uma agência, alguma justificativa deve ser concedida em razão dos altos gastos efetivados. Portanto, o avaliador interno pode equilibrar as críticas com argumentos favoráveis. Méritos, bem como desvantagens, serão apresentados. No entanto, o avaliador interno não pode fazer uma adequada ponderação dos prós e contras. Ele pode hesitar em afirmar expressamente que o projeto é um fracasso. A parcialidade se dá na forma de um equilíbrio enganoso.

O avaliador externo, por outro lado, pode falar com mais franqueza, porque há menos risco de perda de emprego ou antipatia pessoal. Ele é provavelmente melhor em métodos e teorias de avaliação. Esse avaliador externo olhou atentamente para outros programas e pode traçar paralelos que podem ser reveladores. Ele também pode conferir à avaliação a marca da imparcialidade e da alta qualidade.

No entanto, a simples contratação de um avaliador externo não garante uma avaliação objetiva e imparcial. Os avaliadores

externos querem novos contratos. Se eles são muito críticos, eles correm o risco de perder a confiança das pessoas na agência. Assim, eles podem elaborar seu relatório mais para agradar seus contratantes do que para falar a verdade. Os avaliadores externos prostituem suas habilidades (ANDERSON; BALL, 1978, p. 141 et seq.).

Pode-se também defender a avaliação de prestação de contas internas. Na gestão orientada para os resultados, os agentes subordinados devem prestar conta de seus resultados aos seus superiores. A justificativa para isso deve ser que os agentes sabem muito mais sobre o estado de implementação do programa, produtos e resultados do que os diretores. Mas, devido à assimetria de informação, podem trapacear, para o bem ou para o mal. (ver capítulo 8).

9.3 Melhoria formativa

Avaliações para a melhoria da implementação (desenvolvimento, promoção, aprendizado) são mais bem conduzidas domesticamente. Para facilitar o aprendizado de ciclo único (*single-loop learning*), aqueles que precisam da informação devem, de preferência, também acumulá-la. Aprendizado de ciclo único (*single-loop learning*) significa aprender dentro dos limites demarcados por um programa específico ou por uma decisão principal.

Considere o caso do *National Board of Trade* (Conselho Nacional de Comércio Exterior da Suécia)³ fazer uma versão preliminar de um portfólio com informações sobre conservação de energia e divulgá-la para um grupo seletivo de alvos para um teste. Se perguntará aos alvos selecionados se a pasta lhes interessou, se a mensagem é compreensível, se a sua atitude para com a conservação de energia mudou depois de ler as informações, se os seus conhecimentos sobre a conservação de energia aumentaram ou se alguma informação foi omitida. A avaliação deve proporcionar informação que possa ser utilizada para melhorar a pasta.

³ *National Board of Trade* (Kommerskollegium) é uma agência governamental ligada ao Ministério das Relações Exteriores da Suécia e trata do comércio exterior, do mercado interno e da política comercial da Suécia, provendo análises e recomendações ao Governo Sueco desde 1651.

As circunstâncias sugerem inequivocamente que essa avaliação deve ser interna. Na verdade, os autores da pasta devem tanto iniciar, como financiar e realizar a avaliação, a fim de aprender com ela. Uma vez que a pasta representa um ensaio de campo antes de o programa real começar, os administradores pertinentes não têm nada a esconder, o que significa que a pressão contra a objetividade da avaliação interna não tem fundamento. Além disso, as considerações de método são de pouca importância. O que é necessário é a amostragem intencional de informantes interessantes, a fim de extrair informações valiosas através de discussões profundas, sem a realização de delineamentos randomizados complicados e condições controladas.

Uma vez que o grosso do programa tenha sido iniciado, seu monitoramento é necessário. O objetivo ainda é puramente gerencial: para melhorar o funcionamento do programa. O objetivo é investigar se a implementação funciona bem ou se existem barreiras e obstáculos que devem ser superados. Também se procuram gargalos problemáticos, a fim de facilitar e acelerar a implementação. Tudo isso sugere a avaliação formativa realizada internamente por aqueles imediatamente no comando.

O avaliador interno compreende o programa melhor e pode evitar erros devido à ignorância do programa. Ele conhece melhor as pessoas, tendo acesso mais fácil a elas. Ele estará por perto quando a avaliação tiver acabado e, portanto, pode facilitar sua implementação.

Mesmo nesse caso, as objeções ao modelo interno têm pouco peso. A avaliação precisa começar imediatamente e produzir resultados o mais rápido possível. A equipe do programa deve aprender rapidamente com os erros. Objetividade e habilidades em métodos são de menor importância, porque não há decisões rápidas ao alcance dos avaliadores no caso.

A defesa da avaliação interna se torna muito mais forte se a avaliação financeira é feita por auditores externos independentes em processos paralelos separados, independentemente da avaliação interna. (WOHLFARTH, 2014; VEDUNG; DAHLBERG 2013, p. 116 et seq.; VEDUNG, 2017).

9.4 Melhoria aditiva

Se a melhoria aditiva (reconsideração da intervenção) está em jogo, a avaliação deve ser externa. Os avaliadores externos podem trazer novas ideias necessárias imediatamente. No entanto, um argumento pode ser usado a favor da avaliação interna pelo fato de que pode ser recomendável deixar o órgão responsável ter uma palavra a dizer antes de seu programa ser remodelado.

9.5 Conhecimento Básico

As avaliações sob a perspectiva de produção de conhecimento básico devem ser preferencialmente realizadas externamente. O primeiro público em potencial de meta-avaliações, por exemplo, não são agências específicas, mas os operadores governamentais, a comunidade de avaliação, os pesquisadores acadêmicos e os professores no campo das políticas públicas e administração pública. Portanto, a avaliação de conhecimento básico pode ser adequadamente realizada por universidades, institutos de pesquisa e outras instituições de ensino superior. No entanto, as avaliações de conhecimento básico também podem ser executadas internamente. Seria útil para a cultura geral de uma agência reunir e o conhecimento oculto em diferentes avaliações.

OBJETOS DA AVALIAÇÃO

Como o objeto sendo avaliado deve ser descrito (número 4 na Abordagem Oito Problemas de Avaliação)? Em intervenções executadas em múltiplos lugares e com múltiplos *stakeholders*, o objeto de estudo deve ser retratado e apresentado como um caso de algo mais geral. Isso ampliará os horizontes de todos os envolvidos e promove o pensamento paralelo que pode repercutir em outras intervenções. O objetivo é o desenvolvimento de grades conceituais, dentro das quais os aspectos característicos da intervenção podem ser divididos em categorias mais restritas e específicas¹.

Como o alcance das grades é enorme, somente alguns exemplos serão apresentados aqui. A listagem na tabela 10.1 enumera seis rótulos sob os quais alguns objetos de estudo podem ser ordenados e estruturados.

Tabela 10.1
Tipos de Objetos de Avaliação

1. Políticas públicas, programas, componentes de programas, projetos, elementos de projetos;
2. Instrumentos de políticas públicas e pacotes de instrumentos de políticas;
3. Questionários padronizados para coleta de dados;
4. Gestão orientada aos resultados ou aos processos;
5. Políticas setoriais ou orientadas a setores, a áreas e políticas integradas;
6. Cooperação entre agências do setor público (*joined up government*), parcerias, subcontratação.

¹ Ocasionalmente, descrever a intervenção é desnecessário na avaliação. Autoavaliações pessoais, por exemplo, depois de participar de uma conferência ou projeto, podem se concentrar no que foi aprendido sem gastar energia nas descrições dos eventos através dos quais se aprendeu.

10.1 Políticas, programas, componentes, projetos, elementos

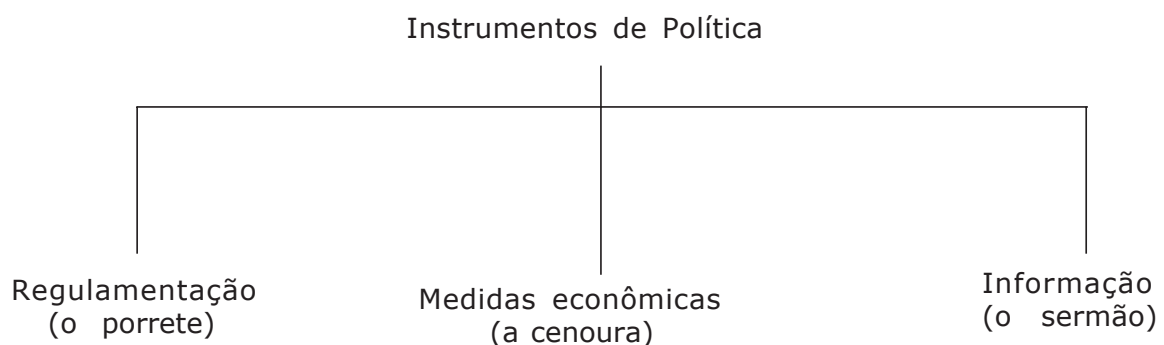
Se a intervenção estudada é uma política, um programa, um componente de programa, um projeto ou um elemento de projeto, pode ser apropriado descrevê-la nos termos de seus problemas, meios e fins. O que é 1) o problema substantivo que jaz na intervenção? Quais são 2) as causas do problema? Quais são 3) as consequências futuras do problema se nada for feito para solucioná-lo? 4) Quais fins (metas) são colocados(as) no mandato da intervenção e como eles são ordenados? 5) Qual amplitude de meios (instrumentos de políticas, etc.) é incorporada na intervenção? A intervenção requer ferramentas regulatórias, econômicas ou informacionais de governança? Passemos a analisar a questão dos instrumentos de políticas de forma mais detalhada.

10.2 Instrumentos de Políticas Públicas e Pacotes de Instrumentos de Políticas

Ao invés de se concentrar, digamos, em políticas ou programas, as avaliações podem focar em instrumentos genéricos de política envolvidos em políticas ou programas. Autoridades governamentais têm à sua disposição três desses instrumentos fundamentais de política, ou da arte de governar (*tools of statecraft*): o porrete, a cenoura e o sermão. Governos podem tanto nos forçar a fazer o que eles querem, nos recompensar ou cobrar materialmente por fazê-lo, ou pregar o que devemos fazer².

Figura 10.1

Três ferramentas de uma Política: Porretes, Cenouras e Sermões



² A classificação tripartite é uma adaptação de um esquema apresentado por Etzioni (1975, p. 5 et seq.) que fala de poder “coercitivo”, “remunerativo” e

O porrete é chamado regulamentação; a cenoura, medidas econômicas (que podem ser incentivos ou desincentivos); e o sermão pode ser nomeado “informação”.

Suponhamos que o governo queira reduzir o consumo de uma *commodity* estrangeira – digamos têxteis chineses. O Estado pode impor uma proibição das importações dessa *commodity* (regulamentação, o porrete), encarecer ou baratear tal *commodity* através de uma taxa (incentivos econômicos, cenoura) ou forçar os importadores a colocar a etiqueta “made in China” na esperança de que essa nova informação reduza as vendas (informação, sermão).

Regulamentação se refere a medidas empreendidas por unidades governamentais para influenciar pessoas por meios de regras verbalmente formuladas, que ditam a seus recipientes a agir de acordo com o que essas regras ordenam. A característica decisiva de uma regulamentação é que a suposta relação de autoridade entre o governante e os governados implica que os últimos são obrigados a obedecer. Expressões como “restrições”, dispositivos de comando e controle ou “diretivas governamentais” são usadas como sinônimos à regulamentação.

Espera-se que uma regulamentação venha acompanhada de ameaças de sanções negativas como multas, prisão ou qualquer outra punição contra os que não acatarem. Se esse for o caso, tanto as prescrições para a ação quanto as sanções estão incluídas no instrumento regulatório. Sanções não são instrumentos específicos além de regulações. Contudo, ameaça de sanções negativas *não* é uma propriedade *definidora* de regulamentação, porque há regulamentações intencionalmente não acopladas a essas ameaças, a assim chamada *lex imperfecta*. O homem com a placa de trânsito

“normativo”. Vedung (1998b) fornece um relato extenso em inglês sobre a classificação alternativa de instrumentos de política com referência à literatura relevante. Pontos de vista interessantes sobre instrumentos de políticas são encontrados em Mickwitz (2006, p. 83 et seq); Hood (1983); e Salamon (1989, p. 30 et seq). Para leitores nórdicos, um panorama das bases de divisão para instrumentos de políticas é fornecida em RRV (1985, pp. 56-58 et seq). Ver também Söderlind e Petersson (1988, p. 213 et seq); Eckhoff (1983); Hermerén e Åström (1982); Vedung, (1982a, p. 19); Klinberg (1980), para livros dinamarqueses em análise de políticas públicas. Para literatura em português ver Ollaik e Medeiros (2011).

vermelha em cruzamentos e esquinas é um desses exemplos: por trás do sinal há uma regra dizendo que somos proibidos de cruzar, mas que os violadores não serão sujeitos a punições³.

Instrumentos de política econômica envolvem ou a distribuição ou a retirada de recursos materiais com a intenção de induzir uma ação ou um estado desejado às coisas. “O Estado dá, o Estado tira”. Ambos os recursos materiais e o conjunto de regras que os acompanha estão incluídos nos instrumentos econômicos. Instrumentos econômicos aumentam ou diminuem o preço de perseguir certas ações. Contudo, os beneficiários das políticas não são obrigados a tomar as ações recomendadas, um fato que torna os instrumentos econômicos muito diferentes das regulamentações. Ferramentas econômicas sempre deixam certa margem aos cidadãos súditos de governança, dentro da qual eles podem escolher por si mesmos se tomarão ou não determinada ação.

Os beneficiários das políticas podem decidir não usar um *incentivo* governamental – um subsídio particular – porque eles podem hesitar em adotar as medidas necessárias para consegui-lo. Ou podem se candidatar a um subsídio porque eles pensam que as demandas que se aplicam a ele são perfeitamente válidas e até mesmo atendem seus interesses. O ponto é: incentivos econômicos não prescrevem nem proíbem as ações envolvidas, mas as tornam menos caras.

³ A delimitação do conceito, contudo, está em forte contraste com diversas definições americanas, as quais são iguais à regulação com intervenção governamental de forma geral. Nos últimos casos, “regulamentação” é uma categoria que abrange tudo, cobrindo todas as formas de controle político. Uma citação do livro de K.J. Meier, **Regulation: Politics, Bureaucracy, and Economics** (1985, p. 1), pode ilustrar esse uso mais amplo: “Regulamentação é qualquer tentativa do governo de controlar o comportamento de cidadãos, corporações ou sub-governos. Em um sentido, regulamentação é nada além de uma tentativa governamental de limitar as escolhas disponíveis aos indivíduos de uma sociedade.” Praticamente tudo que os governos empreendem, então, seria “regulamentação”. Contudo, esse não é o jeito que o termo é usado no contexto atual, onde regulamentação é tida como apenas uma de uma variedade maior de ferramentas que os governos possuem à sua disposição para exercer poder sobre as ações de seus cidadãos.

O mesmo vale para *desincentivos*. Uma taxa arrecadada sobre o tabaco não proíbe fumantes arraigados de deleitarem-se com seus *Ritmeesters* e *Camels*. Os que continuarem a fumar não estão violando a lei; permite-se plenamente a que se entreguem aos seus hábitos, mas o governo tenta desencorajá-los ao fazê-los pagar mais.

Ainda assim, impostos envolvem uma feição coercitiva. Ao realizar a ação na qual o imposto incorre, o agente deve pagar uma taxa, um imposto sobre o consumo de cigarros. Agora, alguém pode se perguntar qual é realmente a diferença entre uma taxa e uma regulação. Ambos parecem envolver uma medida de força autoritária. A diferença é que uma regulamentação proíbe ou prescreve a ação propriamente dita (compra de cigarros), enquanto a obrigação no nosso exemplo de imposto se aplica a outra ação: o pagamento de impostos ao governo. Há uma diferença de princípio entre a proibição da compra de charutos *Ritmeester* ou cigarros *Camels* e a adoção de uma taxa na compra dos mesmos.

Por fim, *informação*, ou “persuasão moral”, cobre as tentativas de influenciar pessoas através da transferência de conhecimento, comunicação de argumento arrazoado, e persuasão. Informação é um termo genérico para campanhas de comunicação, difusão de materiais impressos como brochuras, panfletos, livretos, folders, *flyers*, boletins, panfletos e pôsteres, propaganda, etiquetagem, auditorias, inspeções, programas de demonstração, aconselhamento, conselhos pessoais, programas de treinamento, esforços educacionais e outras formas de aglomeração, embalagem e difusão de conhecimento e recomendações, opções de internet incluídas. Contudo, não se oferece nada além da simples transferência de conhecimento ou raciocínio persuasivo para seduzir – ou desencorajar – as pessoas a fazer o que o governo considera desejável ou indesejável.

Assim como o que ocorre com as ferramentas econômicas, não há obrigação governamental envolvida. Sob circunstância alguma os beneficiários do programa são compelidos a agir da forma delineada na informação. Seguir ou não as recomendações está inteiramente a cargo do público-alvo, porque, por definição, informação não inclui meios de influência mais fortes do que recomendações e raciocínio concomitante. Essa ausência de obrigação faz a informação diferente da regulamentação, que por definição contém regras compulsórias de conduta.

Todavia, informação também é diferente de instrumentos da política econômica, pois não envolve distribuição ou retirada de recursos materiais. A informação distribuída pode muito bem conter argumentos defendendo que os beneficiários do programa de fato sejam agraciados materialmente ao tomar as medidas recomendadas. Contudo, o governo não recompensa materialmente as pessoas que agem, nem priva materialmente as que não agem. As únicas coisas oferecidas são dados, fatos, conhecimento, argumentos e apelos morais.

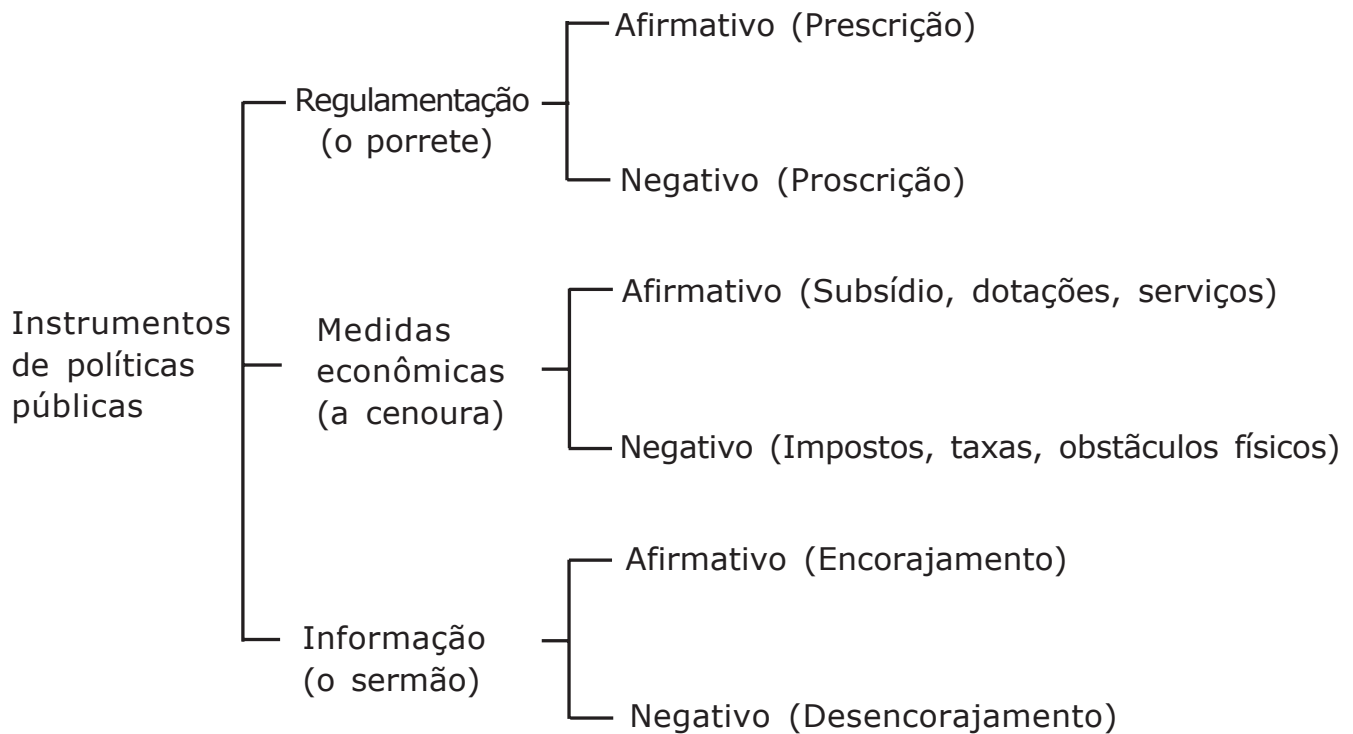
Força derivada da autoridade ou *constrangimento* envolvido na tentativa de governança é a base da divisão (*fundamentum divisionis*) do qual essa classificação tripla procede. Em princípio, a regulamentação é mais coercitiva para os beneficiários das políticas que os meios econômicos; e os instrumentos econômicos são mais coercitivos que a informação. Um proibição da compra de charutos *Ritmeester* impõe mais restrições à liberdade de ação dos fumantes que um imposto adicionado à venda deles, que por sua vez é mais restritivo que informação dissuasiva proclamando que esses meios de prazer sensorial não deveriam ser comprados ou usados.

Uma forma apropriada de ampliar os horizontes dos contratantes da avaliação, assim como fazer a avaliação mais interessante de um ponto de vista das ciências sociais, seria caracterizar a iniciativa avaliada em termos de regulamentação, meios econômicos ou informação⁴. A abordagem dos instrumentos de políticas públicas ao representar a intervenção avaliada parece particularmente render frutos positivos em campos como meio-ambiente, energia, silvicultura, agricultura, políticas regionais, políticas de inovação e transferência de tecnologia (HILDÉN et alli., 2002; MICKWITZ, 2006; ÅSTRAND, 2006; SIMILÄ, 2007; MUNDACA TORO, 2008; STENQVIST, 2013; STERNER; CORIA, 2011).

⁴ Sandahl (1983) aplica tipologia de instrumentos de política para caracterizar a política habitacional sueca. RRV - *Riksrevisionsverket* (1986); Bemelmans-Videc, Rist e Vedung (1998); Hildén et alli (2002); e Mickwitz (2006) discutem a avaliação de instrumentos de política. Na avaliação de informação como um instrumento de política, veja Furubo (1996), Official Publications of Sweden - SOU (1983, p. 34) e Statens Energiverk (1984, p. 2). Aqueles que quiserem ler mais sobre classificações adicionais na estrutura de porretes, cenouras e sermões referir-se a Vedung (1998a, p. 117-121), Vedung (1993a, p. 131-147) e Vedung (2002, p. 94-113).

Todos os três tipos de instrumentos de políticas podem ser fraseados no afirmativo ou no negativo (figura 10.2). Regulamentações podem prescrever ou proibir. Instrumentos econômicos podem distribuir ou retirar. Informação pode persuadir ou dissuadir.

Figura 10.2
Instrumentos de políticas afirmativos-negativos



Instrumentos de políticas públicas também vêm em *pacotes*. Diferenciam-se em pacotes de instrumentos lógicos, horizontais e verticais.

Informação é uma condição necessária para o funcionamento racional de todos os outros controles governamentais. Intervenções públicas requerem que os beneficiários pretendidos sejam informados quanto à existência e ao significado do instrumento de política. Se uma regulação, por exemplo, uma lei estatutária, deve ser efetiva, indivíduos no grupo alvo devem estar ciente de sua existência e de seu conteúdo. Caso contrário, eles não podem acatá-lo racionalmente. Mesmo instrumentos econômicos requerem informação concomitante para funcionar. Beneficiários prospectivos devem saber que já existe um programa de subsídios e sob quais condições o apoio econômico pode ser dado para possibilitar a solicitação para receber subsídios.

Às vezes, o governo também tem que informar às pessoas sobre programas de informação. Depois que novo material de informação for desenvolvido, os usuários em potencial dessa informação são avisados quanto à disponibilidade do material, quanto a ele ser importante e que beneficiários prospectivos devem encomendá-lo. A combinação de instrumentos de política mais a informação sobre esses instrumentos é chamada de *pacote lógico*. Às vezes considera-se necessário avaliar como funciona a informação sobre, por exemplo, uma nova lei.

Nos *pacotes horizontais*, dois ou mais instrumentos são direcionados simultaneamente no mesmo agente envolvido em uma atividade em particular. Por exemplo, um agente pode ser forçado a fazer algo (porrete), mas para facilitar sua ação de acordo com a política (conformidade), o governo pode oferecer auxílio financeiro (cenoura) e conselho (sermão).

Nos *pacotes verticais*, um nível mais alto do governo utiliza instrumentos de política em direção a um nível mais baixo com o intuito de fazê-lo dirigir instrumentos de política a um terceiro nível, ainda mais baixo na cadeia de comando. Por exemplo: um governo (Nível 1) pode usar um tipo de instrumento de política em um agente de implementação de primeira ordem, o agente de implementação (Nível 2) de um novo tipo de instrumento dirigido a um agente de implementação subordinado, e o último (Nível 3) deve fazer um terceiro tipo de instrumento mirado no grupo alvo (Nível 4), que deve fazer alguma coisa.

Regimes de sustentabilidade global como a Convenção de Clima da ONU, incluindo seu Protocolo de Kyoto para combater o efeito estufa, supostamente são implementados através da combinação vertical de instrumentos. A ideia é que a Convenção e seus protocolos (Nível 1) devem ser tomados como um instrumento vinculatório pelos signatários (porrete), para que eles (Nível 2), com base no Protocolo, decidam introduzir instrumentos, que talvez tenham a intenção de atingir os municípios (Nível 3), que por sua vez pretendem atingir os emissores de gases do efeito estufa (Nível 4), que, por sua vez, supostamente devem diminuir as emissões desses gases. A combinação vertical quer dizer que um ou mais intermediários devem estar envolvidos entre o principal tomador de decisões (os países do mundo na Convenção) e os receptores finais pretendidos (os emissores de gases de efeito estufa). Esses intermediários podem pertencer à esfera pública ou podem ser entidades não públicas.

10.3 Questionários Padronizados para Coleta de Dados

Avaliações podem interessar-se em utilizar questionários padronizados para coleta de dados necessários a fim de permitir uma ação apropriada. Tais avaliações são relativas a como esses instrumentos são implementados em instituições específicas e outras localidades, às vezes como a conversa extraída de entrevistados é traduzida em categorias dos questionários pelo entrevistador, e às vezes como os dados assim criados são usados mais além em trabalhos de rotina em instituições e em tomada de decisões. Questionários padronizados existem como documentos, tanto em forma escrita como digital. São usados como adereços em entrevistas face-a-face ou na observação nas premissas relevantes⁵.

Questionários padronizados são frequentemente implementados no setor público com indivíduos, por exemplo, em serviço públicos como cuidado com os idosos, creches infantis e outros programas para as crianças, saúde pública, trabalho social, políticas de emprego, *et cetera*.

O ISV (Índice de Severidade de Vícios) é um método de entrevista estruturada que cobre vício em álcool e drogas, *mais cinco domínios adicionais que são comumente associados com o uso de drogas: médico, legal, emprego, social/família e funcionamento psicológico*. Consiste de um formulário básico com mais de 180 perguntas e um formulário de *follow up* com aproximadamente 150 perguntas. A entrevista básica deve espelhar o período anterior da vida do cliente e as circunstâncias atuais, enquanto o formulário de *follow up* se concentra nas condições atuais. Os fatos capturados serão usados para classificar os entrevistados em várias categorias e tendo assim uma base para decidir quais tratamentos utilizar. Na Suécia, o ISV é aplicado para planejar o tratamento do dependente químico assim como na saúde e no atendimento à saúde. Também é usado na prevenção de crimes e em diversas casas de tratamento e em algumas

⁵ Prior (2003, p. 34; 43 e p. 46 et seq.) delibera sobre as forças e fraquezas particulares de tais *documentos generativos para coleção de dados sobre indivíduos*. Documentos generativos – “documentos que fornecem a estrutura conceitual sobre a qual se relata o mundo” (p. 43).

instituições associadas ao Conselho Nacional de Atendimento Institucional (*Statens institutionsstyrelse*, **Sis**)⁶. Internacionalmente, as entrevistas do ISV são usadas em diversos países europeus e na América do Norte. (<http://www.socialstyrelsen.se/IMS/ASI> per 20071031)

Importantes meios de governo caem fora do esquema porrete-cenoura-sermão. Um caso óbvio é a *organização*. Organizar, reorganizar e desorganizar são princípios cruciais em qualquer política administrativa. Governos que constantemente mudam suas organizações para cobrir novos *assuntos*, *novos problemas*, sinalizam uma ideia de racionalidade, garantem a confiança pública e aumentam a eficácia funcional. Organização é necessária para a provisão de instrumentos de políticas regulatórias, econômicas e informativas. Instrumentos de políticas não podem ser aplicados caso não haja uma organização governamental ou alguém a mando do governo – um mercenário governamental – para fazer o serviço. No contexto presente, ‘organização’ é tratada como uma importante estratégia de governança, mas não como um instrumento de políticas no sentido estrito. Nas seguintes seções, alguns objetos organizacionais de avaliação serão retratados⁷.

⁶ Alexanderson (2006, p. 19 et seq.) é um estudo intrigante de implementação do método ISV. Uma avaliação de como a linguagem comum (da. Fælles Sprog), um esquema de categorias (documento generativo) em forma digital, foi usada para coleta de dados e tomada de decisões na entrega de cuidados com idosos com bases domésticas na Dinamarca. Hansen e Vedung (2005) escrutinizam as teorias de intervenção de meia dúzia de *stakeholders* na Linguagem Comum e suas aplicações.

⁷ Não mencionados aqui estão, por exemplo, modelos divididos Comprador-Provedor, Gerenciamento de Contratos, *Benchmarking* (Pontos de Referência), *Balanced Score Card* (placar equilibrado), Pague-por-performance, Gerenciamento *Lean*, Gerenciamento baseado em valores, Sistemas de gerenciamento de qualidade (e.g. Gerenciamento de Qualidade Total), Comparações Abertas.

10.4 Gestão orientada a resultados ou a processos

Desde a penúltima década do Século XX, governos de todos os níveis se comprometeram à *gestão orientada a resultados*, também conhecida como gestão baseada em objetivos, gestão por objetivos (*Management by Objectives* - *MBO*, em inglês), ou gestão de desempenho. Gestão orientada a resultados não requer o estabelecimento de alguns objetivos e a consequentemente mensuração dos resultados. É uma doutrina de governança muito maior, que consiste em um conjunto de movimentos ou atividades. Na literatura relevante (talvez também na prática), a gestão orientada a resultados é construída, ora em quatro; ora em cinco, seis ou sete; ora em nove ou até mais passos. Na tabela 10.2 a doutrina foi condensada em quatro passos maiores sob a suposição de uma situação organizacional com somente um principal e um agente.

Tabela 10.2

Abordagem de quatro passos para a gestão orientada aos resultados

1. O principal, e.g. uma agência nacional, estabelece objetivos claros e aloca meios à agência como uma soma não especificada.
2. O agente, e.g. um município, clarifica os objetivos, desenvolve indicadores para fazer os objetivos mensuráveis e seleciona meios para atingir os objetivos dentro dos limites orçamentários.
3. O principal anuncia que o monitoramento e a avaliação serão feitos e toma o cuidado para que estas atividades se realizem em uma data posterior.
4. Principal e agente aprendem algo dos resultados da avaliação e do monitoramento, que os instigam a adotar medidas retificadoras.

Estabelecer metas nos *outcomes* é um traço importante, indicado pela expressão “orientado aos resultados” no próprio nome da doutrina de governança. Outra característica importante é a liberdade de ação que agência possui ao escolher meios, métodos e estratégias para alcançar os objetivos. As identificações ou os registros de dados em indicadores de resultados são continuamente coletados através do monitoramento. Depois de um tempo, o principal anuncia que os resultados serão avaliados e que os achados serão

usados para correção e melhora. Essa doutrina de gerenciamento abrangente, que requer governança diretamente nos resultados, difere do gerenciamento orientado aos processos, que foca nos *inputs* (gestão de insumos), atividades internas e *outputs*⁸.

Embora deva ter um importante papel interno na gestão orientada a resultados, a avaliação também pode ser aplicada externamente para aferir um sistema de gestão orientado a resultados como um todo. O gerenciamento orientado a processos também pode ser avaliado (IMAS; RIST, 2009; KUSEK; RIST, 2004; NIELSEN et al 2005; veja também PIHLGREN; SVENSSON, 1989; ROMBACH, 1991; SUNDSTRÖM, 2003, 2006; HERTTING; VEDUNG 2009).

10.5. Orientado a setores, orientado a áreas ou integrado (*mainstreamed*)

Avaliações podem concernir intervenções orientadas a setores ou a áreas, ou intervenções do tipo de políticas integrada. Os Estados de bem-estar social nórdicos foram erguidos sobre princípios setoriais. Barreiras⁹ para o cuidado com a saúde, educação, trabalho, prevenção de crime, agricultura e silvicultura foram criados dentro das fronteiras físicas do Estado-nação e abrangendo todo o território nacional. De tempos em tempos, políticas públicas de um setor inteiro são avaliadas, por exemplo, quando o Serviço Florestal Sueco em 1996-1998 cuidadosamente reviu toda a política florestal da Suécia.

Políticas regionais em todo o mundo são direcionadas a unidades geográficas menores, áreas de apoio dentro dos territórios nacionais. Políticas regionais clássicas eram dirigidas a periferias geográficas privadas do crescimento econômico que caracterizava partes mais centrais do regime. Recentemente, outra política pública orientada a áreas emergiu: iniciativas metropolitanas. A Iniciativa Metropolitana Sueca de 1998-2006 foi orientada à área, pois ela não abrangia o

⁸ Para gestão orientada a resultados ou a processos ver o caso japonês em M. Imai (1996).

⁹ Barreiras (silos em inglês) representam uma parede ou uma barreira colocada por uma organização para manter a mesma focada em obter suas metas e controlar a interferência de pessoas de fora com andamento do programa. Algumas vezes este é um construto organizacional, um insulamento burocrático especificamente desenhado para proteger e servir a hierarquia.

território nacional do reino inteiro, mas sim somente três regiões metropolitanas e dentro dessas, algumas áreas geográficas socialmente vulneráveis nos sete municípios de Estocolmo, Huddinge, Haninge, Södertälje, Botkyrka, Gotemburgo e Malmö. As inúmeras avaliações realizadas sobre a Iniciativa tinha esses programas baseados na área como seus objetos (HERTTING; VEDUNG, 2009).

Objetos sendo avaliados também podem ser detalhados como *iniciativas de política integrada*, às vezes também referidos como intervenções *mainstream*. A ideia é que campos problemáticos, como igualdade de gênero, integração de imigrantes, ambiente de trabalho, política climática (o efeito estufa) e desenvolvimento sustentável não deveriam ser setores em si, mas integrados com todos os setores relevantes existentes na sociedade, integrados com o *mainstream* existente, ou barreiras na sociedade como ela era. Igualdade de gênero, por exemplo, deveria ser uma preocupação para cuidado com a saúde, educação, emprego, serviços sociais, liberdade condicional, agricultura e silvicultura, assim como em outros setores da sociedade. Quando um aspecto como igualdade de gênero é completamente integrado, ele deve cumprir pelo menos duas condições:

- 1) equidade deve ser levada em conta em todos os estágios da formação, adoção e implementação de política em todos os setores de política;
- 2) o item 1) anterior deve ser suplementado por tentativas de agregar efeitos esperados da igualdade de gênero a uma avaliação compreensiva das políticas adotadas e um compromisso para minimizar as contradições entre a igualdade de gênero e as outras políticas. (MICKWITZ, 2006, p. 56 et seq., p. 167 et seq.; MICKWITZ; KIVIMAA, 2007; KIVIMAA; MICKWITZ 2006, KIVIMAA, 2008, p. 19 et seq., p. 55 et seq.; UNDERDAL, 1980; LAFFERTY; HOVDEN, 2003, p. 9; NILSSON; ECKERBERG, 2007).

Enquanto as políticas baseadas na área são trans-setoriais até às comunidades locais, políticas integradas afetam mais de um setor no seu topo, mas iniciando-se nas agências nacionais até os territórios locais é incluída em diversos setores. Outra diferença é que política integrada deve funcionar através do país inteiro, enquanto políticas orientadas à área são restritas a algumas partes geográficas específicas do Estado.

10.6. Governos Conjuntos/Co-Governos, parcerias (governança colaborativa) e subcontratação

Sob esse título podemos discernir:

1) *cooperação* entre duas agências públicas ou mais (conjuntos de governos);

2) *parcerias* entre agências públicas, empresas privadas, organizações da sociedade civil e grupo alvo (governança colaborativa, governança); e

3) *subcontratação* de terceiros pelas organizações do setor público (terceirização).

A cooperação de conjuntos de governos ou cooperação de agência pública implica que agências setoriais devam coordenar seus esforços a fim de resolver os problemas dos indivíduos. Por exemplo, os serviços sociais municipais, a saúde e o atendimento com a saúde do Conselho Municipal e a agência de empregos do Estado nacional deveriam cooperar a fim de reabilitar indivíduos que recebem assistência social, estão doentes ou desempregados. (GOLDSMITH; EGGERS, 2004; TURNBULL, 2003; ver também FORSS, 2007, p.69 et seq., com exemplos; HERTTING, 2003; HERTTING; VEDUNG, 2009: “*Joined-up government around individuals*” no Capítulo 4; e HJORTSJÖ, 2005).

Mais abrangente que cooperação entre conjuntos de governos, as *parcerias* contém não só as agências governamentais, mas também companhias privadas, associações da sociedade civil e até mesmo clientes. Tal cooperação pode acontecer entre a agência de emprego do Estado, os serviços sociais municipais, companhias privadas e associações na sociedade civil para criar empregos (GOLDSMITH; EGGERS, 2004, TURNBULL, 2002; ver também HERTTING, 2003, HJORTSJÖ, 2005). Tanto a parceria (governança colaborativa) e conjuntos de governos se concentram em produzir um resultado integrado e coerente através da formação de um total integrado e coerente.

A *subcontratação* é a contratação de um processo de negócios para uma terceira entidade. Um contrato estipula qual atividade é subcontratada para o terceiro e quanto dinheiro ele receberá para prestá-la. Instituições públicas podem subcontratar certas atividades não essenciais como a operação de restaurantes para servidores, máquinas de café, limpeza de janelas, arranjos florais e manutenção de computadores para o setor privado. A disseminação de informação também pode ser subcontratada para organizações da sociedade civil. Tudo isso pode ser aferido pela avaliação.

MONITORAMENTO

Monitoramento: supervisionar atividades em curso para assegurar que elas se mantenham ativas e sem atrasos para que os objetivos e metas de desempenho sejam realizados.

<http://www.businessdictionary.com/definition/monitoring.html#ixzz2nHpnWsIn>

O monitoramento é uma função contínua na administração pública que se utiliza de uma sistemática coleta de dados de indicadores específicos para prover a gerência e os *stakeholders* de intervenções em andamento com indicação do grau de progresso e conquistas de objetivos substanciais e objetivos no uso do capital alocado (OECD, 2002, p. 27-28 e IMAS; RIST, 2009, p. 16).

O monitoramento pressupõe que a intervenção esteja em andamento. De várias maneiras, *monitoramento* difere de avaliação. Monitoramento engendra coleta de dados de algumas variáveis sem, contudo, relacionar esta atividade à eficácia da intervenção ou às relações de causa-efeito em geral. Em um tipo ideal de monitoramento a coleta de dados é limitada aos resultados brutos enquanto a avaliação pode utilizar estes resultados brutos em esforços para desembaraçar efeitos líquidos. Por esta razão, o monitoramento não é uma avaliação (ver capítulo 7, #6 Problema do Resultado).

Tanto o monitoramento quanto a avaliação giram em torno dos *outcomes*, bem como do *output* e da implementação. Por este motivo não há diferença entre os dois.

Monitoramento é mais uma atividade contínua preocupada com os esforços da governança, enquanto a avaliação é periódica, *ad hoc*, e uma aferição descontínua.

Quando o tipo ideal de monitoramento encontra discrepâncias entre o desempenho real e os objetivos, isto é descrito e relatado, mas não é investigado. O tipo ideal de avaliação, por outro lado, levaria a sério estas discrepâncias e procuraria determinar suas causas, particularmente se elas são devidas a defeitos na intervenção, na implementação, nos *outputs* ou às possíveis contingências no contexto natural.

11.1 A abordagem de 9 passos de monitoramento

O monitoramento pode ser condensado em uma operação de nove passos:

- Passo 1: Articular um tópico geral de monitoramento
- Passo 2: Reconstruir a teoria de intervenção
- Passo 3: Utilizar a teoria da intervenção para determinar Onde e O Que monitorar
- Passo 4: Processar o tópico de monitoramento em uma questão de monitoramento abrangente
- Passo 5: Detalhar a questão de monitoramento abrangente em subquestões específicas
- Passo 6: Coletar dados
- Passo 7: Analisar os dados
- Passo 8: Valorar as descobertas
- Passo 9: Relatar as descobertas

Os nove passos estão listados em uma fraca ordem lógica. Um tópico geral de monitoramento deve ser formulado antes que uma teoria de intervenção pertinente possa ser reconstruída ou uma pré-existente seja selecionada. Questões de monitoramento devem ser redigidas e delimitadas antes que uma escolha sensata de métodos de coleta de dados seja realizada, os dados devem ser coletados antes que a análise deles possa ser feita e a análise de dados deve preceder o relato das descobertas. Para elucidar cada passo, utiliza-se a Declaração de Combustível como exemplo. (Figura 3.4 no capítulo 3).

Passo 1 - Articular um tópico geral de monitoramento

Frequentemente, o ponto de partida prático para o monitoramento é um tópico disforme como “faz nove meses desde que a Declaração de Combustível entrou em vigor; é hora de fazer uma investigação; devemos descobrir como a implementação procede; a conformidade pode estar fraca em alguns lugares, então temos que tomar certas medidas”. E, como uma continuação, os contratantes podem perguntar aos monitores: “Vocês podem, por favor, nos dar algumas ideias sobre o que nós devemos prestar mais atenção?”

Passo 2 - Reconstruir a teoria de intervenção

O bom monitoramento deve ser baseado em uma teoria de intervenção cuidadosamente reconstruída. Isso implica no devido esclarecimento e refinamento dos pressupostos que lidam com o que a intervenção é feita para atingir e como esta conquista deverá acontecer. A teoria de intervenção reconstruída, concebida como um diagrama mostrando uma corrente de implementação (ver figura 3.4), deve fornecer a orientação quanto o monitor decide em que estágios as verificações empíricas serão realizadas. A ideia é que a implementação perfeita baseia-se em um número de pré-requisitos que devem ser atendidos para que a intervenção seja eficaz. É a tarefa do monitoramento apontar estes pré-requisitos e investigar se eles foram realizados.

Passo 3 - Utilizar a teoria da intervenção para determinar Onde e O Que monitorar

O monitoramento pode atender a todos os componentes da cadeia de execução. O ponto onde se espera que o *output* aconteça é crucial para programas inaugurados recentemente. No caso da Declaração de Combustível, é nas concessionárias de veículos onde a informação exigida pelo governo deve ser exibida. Um pré-requisito óbvio para o sistema de governança funcionar como o esperado é que os compradores de carro possam ser expostos à informação. Assim, os materiais informativos exigidos pelo governo devem ser exibidos nos estabelecimentos de venda de carros. Para esclarecer:

Pré-requisito 1: Para ter um efeito na compra de carros e no uso de combustíveis, a informação exigida pela Declaração de Combustíveis devem estar exibidas nos locais de venda.

Agora, suponha que a equipe de monitoramento decida verificar em que nível este pré-requisito é cumprido (O Quê?) nas várias concessionárias de veículos (Onde?).

Passo 4 - Processar o tópico de monitoramento em uma questão de monitoramento abrangente

No 4º passo, o pré-requisito 1 do passo número 3 é reformulado na primeira questão de monitoramento. Desta forma:

Questão de monitoramento 1: As informações exigidas pela Declaração de Combustível estão em exposição nos locais de venda?

Note que deve haver uma simetria total entre o pré-requisito e a questão de monitoramento. Note também que a questão exige uma descrição: As informações ordenadas estão em exposição nos locais de venda?

Passo 5 - Detalhar a questão de monitoramento abrangente em subquestões específicas

Muitas vezes, intervenções em seus pontos de entrega podem requerer uma enérgica operacionalização em indicadores para poder fornecer orientações para o monitoramento (BRYMAN, 2008, p.143 et seq.). Na verdade, a criação de indicadores é uma das complexidades do monitoramento e da avaliação. Aqui pode ser pertinente lembrar ao leitor das dificuldades envolvidas na operacionalização, como por exemplo, conceitos como a qualidade do cuidado com as crianças, saídas de livros da biblioteca, ou pesquisas e inovações científicas.

No entanto, a Declaração de Combustível é extraordinariamente fácil de operacionalizar. Bem definidos em documentos emitidos pela Agência do Consumidor, o programa exige que todas as concessionárias coloquem uma etiqueta (*etikett*, em sueco) no para-brisa de cada carro novo, coloquem um cartaz (*affisch*) na parede, e tenham um livreto (*broschyr*) colocado em um local visível. O conteúdo destes três meios de informação também é bem definido e delimitado. A partir disto, pode-se fazer três operacionalizações do pré-requisito 1 no 3º passo:

Pré-requisito 1a: Informações escritas sobre o consumo de combustível a cada 100 km (quilômetros) devem ser fornecidas através de uma etiqueta no para-brisa de cada carro novo no local.

Pré-requisito 1b: Um cartaz deve ser afixado à parede, mostrando informações escritas sobre o consumo de combustível a cada 100 km para todos os novos modelos de carros vendidos pela concessionária.

Pré-requisito 1c: Um livreto deve ser colocado em algum local visível, fornecendo informações sobre o consumo de combustível a cada 100 km para cada novo modelo de automóvel comercializado em todo o país.

Estes pré-requisitos devem, então, ser reformulados em subquestões específicas dentro da questão abrangente de monitoramento. É claro que deve haver simetria entre os pré-requisitos na teoria da intervenção e as subquestões de monitoramento. Desta forma:

Subquestão específica 1a. Está, a informação relativa ao consumo por 100 km, fornecida por escrito através de uma etiqueta no para-brisa de cada carro novo nos salões da concessionária?

Subquestão específica 1b. Há um pôster afixado na parede informando por escrito sobre o consumo de combustível a cada 100 km para todos os novos modelos de carros vendidos pela concessionária?

Subquestão específica 1c. Existem livretos colocados de maneira visível, fornecendo informações sobre o consumo de combustível por 100 km para cada novo modelo de automóvel comercializado em todo o país?

Passo 6 - Coletar dados

A coleta de dados deve ser regida pelo monitoramento específico das subquestões 1a, 1b e 1c. Nesse exemplo, deve-se descobrir se a informação sobre o consumo de combustível exigida está em exposição, nos lugares corretos e com o conteúdo correspondente ao ordenado.

Como será extensivamente enfatizado no capítulo 12, há quatro e apenas quatro grandes métodos de coleta de evidências: documentação, entrevistas, observação e audiovisual. Aqui nós vamos supor que o grupo de monitoramento escolheu a abordagem interrogativa, ou mais especificamente o uso de questionário com respostas escritas como método de coleta de dados. Mil questionários são enviados. Supondo que 750 concessionárias preencheram os questionários e os enviaram de volta ao grupo de monitoramento.

Passo 7 – Análise de dados

Como 7º passo, os monitores realizarão a análise dos dados de resposta espalhados nos 750 questionários para criar um quadro geral compreensível e conciso. Eles devem usar algum sistema classificatório em que os dados serão significativamente ordenados.

Vamos supor que a análise divide o material em seis categorias: concessionárias de veículos que exibiram 1) todos os três tipos de informação, 2) dois tipos de informação, 3) um tipo de informação, 4) nenhum dos tipos de informação, 5) desconhecido, e 6) dados inexistentes. Além disso, sugerimos que as concessionárias que exibiram dois tipos e um tipo de informações, respectivamente, sejam classificadas de acordo com o tipo exibido. Quantos mostraram a etiqueta e o cartaz, etiqueta e o livreto, e o pôster e o livreto? Quantos mostraram apenas a etiqueta, apenas o cartaz ou apenas o livreto? E assim por diante.

Os monitores também podem começar a se perguntar sobre as razões para as variações de exibição. Por exemplo, há alguma relação entre variação e:

- a) a ignorância das exigências da Declaração de Combustível (compreensão)?
- b) a falta de vontade (vontade), por exemplo, que alguns vendedores temiam consequências econômicas negativas da exposição enquanto outros esperavam consequências positivas?
- c) que nenhum material de informação estava disponível para exibição (capacidade)?

Na busca por explicações, os monitores deixam o monitoramento e entram na avaliação formativa. Não há nada de errado com isto. Monitoramento e avaliação podem ser combinados no mesmo projeto. Entretanto, buscar explicações é diferente; pertence à avaliação, não ao monitoramento.

Passo 8 – Valorar as descobertas

O passo número 8 envolve a aplicação de critérios de méritos para os resultados empíricos. Na abordagem de acompanhamento recomendada aqui, os critérios de valor embutidos na própria intervenção são os que se aplicam. Em outras palavras, o monitoramento se inscreve em uma teoria descritiva de valoração (ver Capítulo 5).

Para exemplificar, nós escolhemos o Pré-requisito 1: Para haver efeitos na compra de carros e no consumo do combustível, a informação exigida pela Declaração de Combustíveis deve ser exibida

nos locais de venda. As dimensões de critério são etiquetas, cartazes, e livretos sobre o consumo de combustível dos novos carros exibidos com as concessionárias.

O padrão de desempenho aceitável nestes critérios de mérito é 100% de realização do objetivo. A Declaração de Combustíveis obriga todas as concessionárias a fixar etiquetas nos para-brisas, fixar cartazes nas paredes, e exibir o livreto no local. Assim, o desempenho real nos três critérios deve ser comparado com o padrão de desempenho de 100%. Outra abordagem para a seleção de padrões seria avaliação comparativa (*benchmarking*): escolher o melhor registro encontrado com qualquer revendedor no estudo como padrão contra qual a avaliação será medida.

Para ilustrar, voltemos às 700 em 1000 respostas aos questionários. Em 100 casos, 80% da informação estava corretamente exibida; em 200, 70%; em 300, 60%; e em 1000 casos, 50%. Como padrão de desempenho, neste caso, podemos escolher a taxa de 90% como padrão de referência, à luz do que o desempenho de 100 casos é bom, 200 casos são decentes, 300 casos são aceitáveis, e todos os 1000 casos, menos satisfatórios. Outro *benchmark* seria a pontuação média de todos os revendedores investigados.

Em monitoramento, pontuações analíticas são usualmente empregadas, ou seja, valores são atribuídos a cada parte do objeto de avaliação sem alocar um valor único e universal. Em nosso caso da Declaração de Combustíveis, pontuação analítica implicaria que as etiquetas, os cartazes, e os livretos receberiam notas separadas. O pré-requisito das etiquetas pode ser classificado em satisfatório, o requisito do cartaz razoável, e o requisito do livreto poderia ser insatisfatório, por exemplo, em relação ao padrão selecionado. Unir as diferentes pontuações em um valor único para o desempenho geral do objeto de avaliação (*pontuação global*) pode ser deixada para os tomadores de decisão.

Passo 9 – Relatar as descobertas

Evidentemente, os resultados do monitoramento devem ser narrados aos membros do conselho, operadores do programa, e outros potenciais *stakeholders*. De preferência, deve-se relatar (Passo 9) de tempos em tempos durante o processo de monitoramento, o que não descarta abrangentes relatórios finais. Se os resultados se

mostram satisfatórios, nada aconteceria. Se eles não parecem tão bons, a agência pode lançar uma avaliação para descobrir as causas. Se eles parecem muito ruins, medidas concretas podem ser criadas imediatamente para melhorar a implementação.

Tendo proporcionado este panorama de 9 passos de monitoramento retorna-se agora à avaliação, começando com avaliação de impacto e pensamento geral de causa e efeito. Antes disso, será apresentada uma breve exposição dos métodos de coleta de dados.

MÉTODOS DE COLETA DE DADOS: QUATRO E APENAS QUATRO

Em avaliação de intervenções públicas, há quatro e apenas quatro amplas formas de coleta de dados: através de documentação, meios audiovisuais, interrogatórios, e métodos de observação (CRESSWELL, 2013, p.145 et seq; também LAUNSON; RIEPER, 2005, p.109 et seq; MATHIESEN, 2004, p. 37 et seq., p. 48). *Quatro e apenas quatro*. Quanto a isso, a avaliação (e monitoramento) não se diferencia das ciências sociais em geral. Com tais subgrupos, obtemos o seguinte quarteto de abordagens.

- | | |
|---------------------------------|--|
| 1 Métodos documentais | Textos, estatísticas, registros, atas |
| 2 Métodos audiovisuais | Fitas de vídeo, fotografias, desenhos, mapas |
| 3 Métodos interrogativos | Entrevistas (individuais-grupos de interesse (<i>grupos focais</i>),
Questionários (postal-digital; respostas por escrito ou digitais), relatos solicitados pela avaliação (diários, histórias) |
| 4 Métodos de observação | Visitas <i>in loco</i> ; anunciado-não anunciado; oculto-aberto; participante-não participante; avaliador em seu papel formal-avaliador apresentado como cliente |

12.1 Métodos Documentais

Os *métodos documentais* envolvem o uso de dados em forma textual ou numérica produzidos pelo próprio processo para prover informações. Textos escritos, cartas em papel ou digitais, atas de reuniões, protocolos, consultas públicas, registros de clientes e estatísticas pertencem a este conjunto. Métodos documentais são discretos, indicando que a informação não foi gerada a pedido do

avaliador/pesquisador especificamente para o estudo em mãos, mas sim para outros propósitos, intrínsecos ao assunto em foco. Os métodos documentais se interessam por dados disponíveis de outros estudos, esperando para ser analisados e interpretados. Quando produzidos agentes de intervenção, populações alvo e outros informantes não tinham conhecimento do futuro uso dos documentos em outros contextos. Métodos não intrusivos são imunes aos *efeitos de mensuração reativa* (ver capítulo 13), ou seja, o conteúdo dos documentos não está contaminado pelo fato dos produtores saberem que estavam sendo monitorados ou avaliados¹.

Considere os registros dos clientes! Habitualmente, as equipes de intervenção coletam uma boa quantidade de informação sobre os participantes da intervenção. Tais registros, seja em papel ou em formato eletrônico, podem variar de tortuosas narrativas a lacônicos *bits* (pedaços) de inteligência organizados de acordo com um sistema de categorias pré-estruturadas. Tanto os regimes de tratamento de saúde e de atendimento médico de pacientes em um hospital são meticulosamente registrados. Bibliotecas acompanham os nomes, coordenadas, livros emprestados e livros retornados dos mutuários. No exemplo da Declaração de Combustível, contratos de compra entre empresas entregando carros novos e distribuidoras são escritos e catalogados, assim como os contratos entre concessionárias e compradores. Todas estas são amostras de registros não produzidos a mando dos avaliadores, mas que podem ser utilizados como fontes documentais de dados pelos avaliadores.

12.2 Métodos Audiovisuais

Novas formas de dados e de métodos de coleta de dados emergem continuamente na literatura. *Métodos audiovisuais* é uma família de tais métodos; eles não foram incluídos como uma categoria separada na edição inglesa deste livro. Dados visuais e de áudio pertencem a esta categoria. Fotos, desenhos e mapas, digitais ou em papel, são incluídos, tal qual gravações como programas de rádio ou combinações como seminários filmados e apresentações em

¹ Maiores informações sobre métodos documentais em Bryman, 2008, p. 514-536.

conferências são casos em questão. Como os métodos de observação em forma textual ou numérica, eles não são gerados pelo avaliador para a avaliação em curso. (CRESSWELL, 2013, p. 145 et seq, p. 160 et seq).

12.3 Métodos Interrogativos

Os *métodos interrogativos* extraem respostas diretamente das pessoas através do uso de questões. Uma interrogativa é “uma sentença de pesquisa que pede por uma resposta”. Os métodos interrogativos obtêm pedaços de informações feitos sob medida para a investigação em mãos, em contraste com os métodos documentais que se baseiam em dados produzidos para outros propósitos. Esta é a principal linha divisória entre os métodos interrogativos e documentais.

Os métodos interrogativos incluem entrevistas e o uso de questionários (levantamentos). Em entrevistas a informação é coletada através de conversas entre os entrevistadores e os entrevistados pessoalmente ou via telefone ou Skype. (BRYMAN, 2008, p. 191-213, p. 436-471; HAMMOND; WELLINGTON, 2013, p. 11 et seq). Na entrevista semiestruturada o entrevistador utiliza “uma breve lista de pontos de memória (itens a lembrar) de áreas a serem cobertas” (BRYMAN, 2008, p. 695). As entrevistas também são realizadas nos grupos de interesse - os chamados grupos focais (BRYMAN, 2008, p. 472-491). *Grupos focais* é um tipo de entrevista onde há vários participantes além do avaliador/facilitador/moderador, estando todos eles reunidos em um único local. Há uma ênfase em questionar um tópico restrito, em registrar a interação entre os participantes do grupo e construir uma resposta conjunta (BRYMAN, 2008, p.694). Em ambas as entrevistas, individual e coletiva, o avaliador pode usar papel e caneta, ou equipamentos como gravadores, câmeras e computadores.

A coleta de dados através de *questionários* implica na reunião de informações “em mais de um caso (usualmente muito mais que um) e em um único ponto temporal de forma a coletar um corpo de dados quantitativos ou qualitativos em conexão com duas ou mais variáveis (geralmente muito mais que duas), que serão, então, examinados para detectar padrões de associação”. Os dados são, predominantemente, coletados através de questionários de autopreenchimento onde o formulário é estruturado com ambas as

questões e as alternativas das respostas especificadas com antecedência. Ocasionalmente os questionários são preenchidos com o entrevistador presente para esclarecer itens que os informantes achem confusos. Nestes casos, o entrevistador (ou seus ajudantes) recolhe os questionários quando preenchidos. Em um tipo ideal de autopreenchimento de questionários, o entrevistador não está presente, os formulários são enviados por correio (e-mail) e depois de preenchidos são enviados de volta pelo correio (e-mail) pelos entrevistados. O uso de questionários como instrumentos de coleta de dados também é chamado de pesquisa de opinião ou algo semelhante (BRYMAN, 2008, p. 215-252, p. 693, p. 699).

A utilização de *diários* solicitados pela avaliação é outra forma de método interrogativo. Bryman os considera como questionários de autopreenchimento (2008, p. 224-229) porque os diaristas são orientados a utilizar os moldes criados pelo avaliador em seus diários. Diários podem ser resumidos a registros onde os diaristas, em termos gerais, relatam o que eles estão fazendo em diferentes etapas. Diários também podem ter descrições expandidas ou relatos detalhados sobre práticas eficazes ou falhas. Por outro lado, o uso de diários já existentes, relatos, histórias, narrativas, etc, produzidos espontaneamente são exemplos de métodos documentais.

Pelo fato da equipe de intervenção e dos clientes da intervenção estarem cientes de que as respostas por eles fornecidas serão utilizadas para propósitos de avaliação, suas revelações não estão imunes aos efeitos de mensuração reativa.

Conversar com pessoas envolvidas e extrair seus relatos é muito importante para os avaliadores. Em nosso caso da Declaração de Combustível, avaliadores podem querer saber se os potenciais compradores de carros possuem ciência das informações e compreendem seu significado. Para este fim, eles, presumivelmente, entrevistariam uma amostra de compradores de carros ou pediriam a eles para responderem a questionários. Concordamos com Carol Weiss (1972a, p. 53 et seq) em que muito da pesquisa de avaliação depende de entrevistas e questionários, porque a informação necessária sobre participantes da intervenção é difícil de ser adquirida de outra forma. O avaliador pode também entrevistar vendedores ou, talvez, outra categoria de pessoas das concessionárias nos serviços de atendimento ao cliente (ao consumidor) e que fazem com que os clientes compreendam a informação disponível.

12.4 Métodos de Observação

Existem deficiências nos métodos interrogativos. Mais problemáticas, talvez, são as lacunas entre o que as pessoas dizem fazer e o que elas realmente fazem. Observar o comportamento das pessoas diretamente parece ser a solução óbvia para esta lacuna. *Métodos de observação* – nosso quarto conjunto – implicam em reunião de dados através do uso dos sentidos do próprio avaliador. Ele realiza visitas aos locais para ver com seus próprios olhos o que está acontecendo e escuta, com seus próprios ouvidos, o que pessoas pertinentes estão dizendo. Ele pode utilizar equipamentos técnicos como gravadores, câmeras, papel e canetas. Ele pode utilizar fotografias de aviões, e até de satélites.

Preferivelmente, os dados devem ser recolhidos sistematicamente. O procedimento menos estruturado é fazer anotações do que é observado. Uma forma melhor é, talvez, começar com um conjunto de questões que serão metodicamente investigadas enquanto mantém-se a mente aberta para novas questões que podem surgir durante as sessões de observação. O produto de uma observação pode gerar outras observações, em uma espécie de bola de neve.

A força da observação direta reside em sua confiabilidade. Além disso, produz uma percepção mais profunda do que entrevistas e, particularmente, questionários, porque a observação das ideias e ações daqueles interessados pode se estender por um período de tempo maior. Por outro lado, o método é demorado e caro.

Vejamos a Declaração de Combustível novamente. A verificação da conformidade dos revendedores de automóveis com os requisitos de informação – fixando etiquetas a para-brisas, pendurando pôsteres nas paredes e distribuindo folhetos em alguns lugares nas instalações da concessionária – pode ser feita através de visitas *in loco* e observação direta. Os avaliadores podem visitar as concessionárias, observar e fazer anotações.

Algumas vezes, é importante para os avaliadores sair de seus escritórios e bibliotecas para fazer observações *in loco* pessoalmente.

Uma questão controversa é se as visitas *in loco* devem ou não ser *anunciadas*. As pessoas envolvidas devem ser avisadas previamente sobre o lugar, datas e horários da sessão de observação?

Ou o avaliador deve apenas esgueirar-se sem avisar ninguém? A filosofia por trás da divulgação é dar aos implementadores uma chance de corrigir eventuais deficiências, antes da avaliação realmente começar. Esta filosofia está muito bem alinhada com o propósito de aperfeiçoamento da avaliação.

Uma questão concomitante é se a sessão de observação deve ser conduzida abertamente (“Eu sou o avaliador da *National Consumer Agency*”). Então ficaria claro a aqueles interessados com quem eles estariam lidando. A observação também poderia ser oculta, como por exemplo, por trás de um espelho falso e opaco, o que asseguraria uma observação distante e independente.

A *observação participante* é diferente da sua *contraparte não participante*, embora esta distinção não seja sempre feita nos livros-texto de métodos de pesquisa (BRYMAN, 2008, p. 253-272). Na observação participante, o avaliador e sua equipe são mais do que observadores fazendo ocasionais visitas *in loco*. Eles observam e escutam enquanto participam da atividade a ser avaliada. Isso implica em uma grande imersão no ambiente social onde o avaliador deseja descobrir o sentido das ações dos envolvidos.

Um tipo de observação participante está intimamente associado ao sociólogo alemão Günter Wallraff. Ao se disfarçar e fingir ser turco durante vários anos, Wallraff utilizou observação participativa para verificar como seus concidadãos da Alemanha Ocidental responderiam aos trabalhadores turcos convidados².

No caso da Declaração de Combustíveis, este tipo de observação participante implicaria que os observadores – ou aquele que atuem em seu favor – fingiriam ser supostos compradores de carros a fim de registrar como as informações são tratadas pelas concessionárias e pelos vendedores. No caso da Ajuda aos Refugiados Afegãos os investigadores fingiriam ser refugiados afegãos pedindo por barracas e cobertores para os operadores do Crescente Vermelho no campo de refugiados.

Por se basear no engano, a observação participante no estilo de Wallraff é inaceitável na avaliação do setor público.

² Os métodos investigativos de Wallraff levaram à criação do verbo em sueco *wallraffa* que tem por significado “expor uma má conduta interna por assumir um papel” que foi oficialmente incluído na lista de palavras da *Swedish Academy*.

O uso de *fotografias*, tiradas pelo próprio avaliador ou por seus associados para o propósito avaliativo, são certamente observacionais³. Por outro lado, a utilização intensa de fotografias não solicitadas tiradas por terceiros deve ser considerado como um método audiovisual⁴. Na pesquisa de avaliação, fotografias são utilizadas para se fazer avaliações de complicados efeitos de políticas sobre o uso da terra, florestas, pântanos, desenvolvimento urbano, e assim por diante. (COLLIER, J; COLLIER, M, 1986; WAGNER, 1979).

As fotografias podem ser utilizadas no exemplo da Declaração de Combustíveis para mapear os materiais de informação com os vendedores de carros, como, por exemplo, os pôsteres estão pendurados e como os folhetos estão colocados. No Programa de Ajuda aos Refugiados Afegãos, imagens para a avaliação podem ser tiradas com relação a várias características dos campos de refugiados no Paquistão.

Ocasionalmente, alguma reflexão é necessária para se traçar a linha que divide os métodos de observação e métodos documentais. Como foi visto, observadores podem desempenhar um papel formal no sistema a ser monitorado. Como um detento da prisão de Oslo, Johan Galtung⁵ fez nítidas observações sobre a vida na penitenciária e das atitudes dos detentos acerca da vida na prisão. Depois, ele escreveu um livro sobre sua estadia ("Fengselssamfunnet", 1959). No sistema da Declaração de Combustíveis, isto corresponderia a um comprador de carro registrando suas observações em um diário sobre como as três peças de informação são tratadas nas concessionárias visitadas sem ser pedido pelo avaliador. Usar esta informação em uma futura avaliação seria um caso de método de documentação e não de observação. Se o diarista agiu ao comando do avaliador e o diário é, posteriormente, utilizado como fonte de dados na avaliação, isto seria um caso de método interrogativo.

Ao invés de se basear em apenas um método de coleta de dados, recomenda-se a triangulação. Bryman (2004, p. 700) define triangulação como "o uso de mais de um método ou fonte de dados

³ Maiores informações sobre fotografias, ver Bryman (2008, p. 424-434).

⁴ Bryman (2008, p.519).

⁵ Sociólogo e matemático, nascido em 1930 na Noruega. Considerado como o mais importante pioneiro nos Estudos sobre Paz e Conflitos.

no estudo de um fenômeno social, de modo que os resultados possam ser cruzados” (Também BREWER; HUNTER, 1989; PATTON, 1987, 1990; GUBA; LINCOLN, 1981, 1989; KOSECOFF; FINK, 1982; HAMMOND; WELLINGTON, 2013, p. 145 et seq).

Retornando ao caso da Declaração de Combustíveis, os avaliadores podem enviar questionários para uma grande amostra de pontos de vendas de carros sobre os locais de colocação de rótulos, pôsteres e folhetos em suas dependências. Isto poderia ser combinado com visitas do avaliador a um conjunto menor de concessionárias para tirar algumas fotos do material de informação e para observar e fazer anotações. As imagens, assim como as anotações, podem ser utilizadas como um incentivo para atrair os vendedores e consumidores a falar. Os métodos de coleta de dados devem ser triangulados para esclarecer as mesmas questões avaliativas.

EFEITOS: EXPERIMENTAÇÃO EM DOIS GRUPOS

Pesquisa de avaliação é, primeiro e mais importante, um processo de aplicação de procedimentos científicos para acumular evidência confiável e válida sobre de que modo e até que ponto atividades específicas produzem efeitos ou *outcomes* particulares.

Leonard Rutman, *Evaluation Research Methods* (1977)

A lei da causalidade, como tanta coisa que ainda vigora entre os filósofos, é uma relíquia de épocas idas, que sobrevive, como a monarquia, apenas porque se supõe, erradamente, ser inofensiva.

Bertrand Russell, *Mysticism and Logic* (1918)¹

A avaliação de forma geral é frequentemente vista como um procedimento de pesquisa acadêmica para testar o impacto de uma intervenção. W. Edwards Deming, guru do movimento de controle de qualidade, disse em um comentário muitas vezes citado – “Avaliação é o estudo de causas”; David Nachmias (1979, p. 9) – “No coração de todas as atividades de pesquisa e de avaliação de políticas está a ideia de causalidade”; e o filósofo sueco Tore Nilstun – “No contexto da avaliação de reformas regulatórias, a análise causal é uma necessidade.”

Contra isso, Scriven (1991, p. 137) discutiu: “A avaliação às vezes descobrirá a causa do sucesso ou do fracasso, do mérito ou da incompetência – mas nem sempre; não é seu dever, e tampouco deveria ser seu objetivo... a tarefa-chave para o avaliador é avaliar... Não é a tarefa do avaliador determinar a [causação de uma condição].”

¹ Em português, citação pode ser encontrada em RUSSELL, Bertrand. Sobre a Noção de Causa. In **Misticismo e Lógica**. Tradução de Wilson Velloso. São Paulo: Companhia Editora Nacional, 1957, p. 204.

Pode-se discordar de Scriven que a tarefa do avaliador não é determinar a causação da intervenção (SHADISH Jr.; COOK; LEVITON, 1991, p. 73, p. 114). Tal aferição de impacto é potencialmente importante, particularmente por tomadores de decisão nos níveis mais altos. E é enfatizada pelo atual movimento pelas políticas públicas baseadas em evidências.

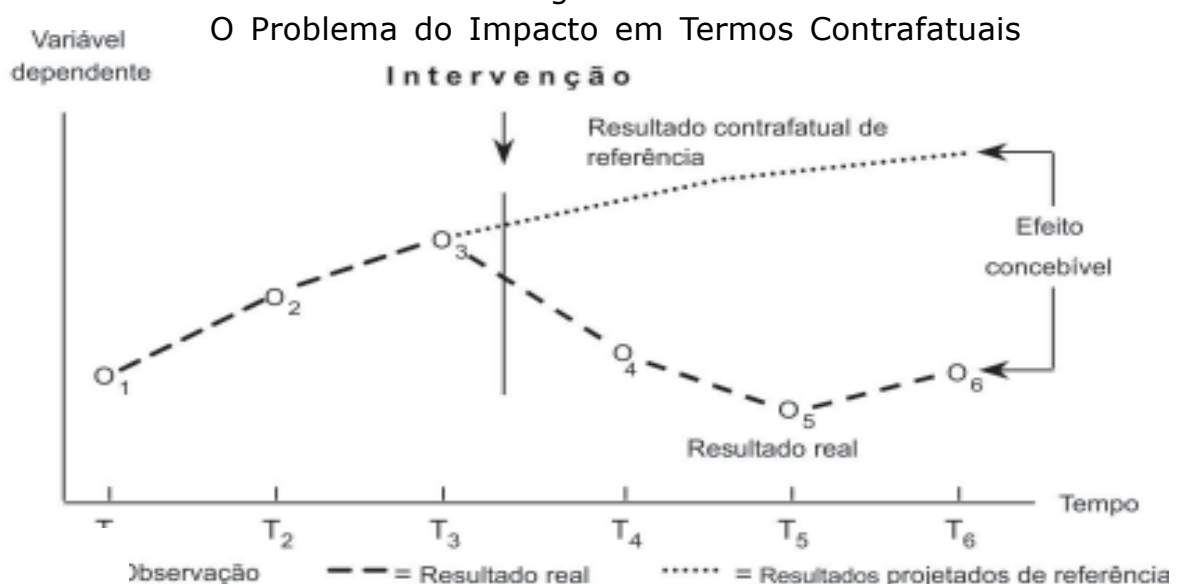
Por outro lado, a avaliação não deve ser igual à aferição de impacto. Na aferição de serviços do governo como segurança pública, transporte de lixo, utilidades públicas, parques e recreação, transporte público, serviços para idosos, bibliotecas e museus, pede-se que cidadãos e clientes avaliem o mérito do que eles recebem. Geralmente não há questão de causalidade envolvida. Contudo, tanto essa abordagem não causal, como outras similares, são ferramentas de avaliação imprescindíveis para gerenciamento e direção.

A aferição de impacto se endereça aos problemas 5 e 6 na Abordagem Oito Problemas da Avaliação (OPA) e particularmente 6.2 e também 5.2. A complexidade fundamental é desemaranhar os efeitos da intervenção dos efeitos que não são da intervenção. *Outcomes* podem ser contingentes sobre os efeitos que não são da intervenção, ocorrendo simultaneamente com ela. De forma geral, esse problema de impacto é extraordinariamente difícil de resolver.

Compreender o *modus operandi* presumido da análise causal clássica não é particularmente trabalhoso. O resultado de fato da intervenção instalada é comparado a um resultado de referência (*reference outcome*) mostrando o que o resultado teria sido caso não houvesse uma intervenção. Se todas as outras forças motrizes causais fossem idênticas em ambas as situações, a diferença subsequente deve ser atribuída à intervenção.

Expôr o problema numa forma contrafactual, como visto na figura 13.1, traz mais esclarecimentos.

Figura 13.1



Novamente, busca-se saber qual teria sido o resultado caso não houvesse um programa, e comparar isso com a hipótese de que houvesse um programa. Para separar os efeitos programáticos dos não-programáticos, comparou-se o resultado que de fato ocorreu com o resultado que ocorreria caso o programa não tivesse sido instituído. Como o último resultado nunca ocorreu, ele vai *contra os fatos*; é *contrafactual*. A diferença do resultado real para o resultado contrafactual constituiria os efeitos do programa.

Um pré-requisito central da análise de impacto é gerar uma série temporal de dados no resultado de fato *antes* da intervenção (*linha de base*) e por algum tempo *depois* da intervenção. O aspecto histórico da questão do impacto é crucial e deve ser fortemente enfatizado.

É esclarecedor ter em mente que tanto os efeitos quanto as causas são *diferenças*. Um efeito é uma diferença entre o resultado de fato e o resultado que teria ocorrido caso não houvesse nenhum programa. A causa é a discrepância entre ter e não ter um programa. (NILSTUN; HERMERÉN, 1984, p. 130 et seq.)².

Pode-se argumentar que aferições de impacto realizadas de forma neutra são importantes para a vida política porque os efeitos da intervenção são fortemente politizados. Legisladores, gestores de programas e grupos de interesse a favor da intervenção estão inclinados a reivindicar crédito por supostos sucessos e desviar da culpa por supostos fracassos. Todos os aspectos positivos que acontecem posteriormente são creditados como efeitos do programa, enquanto consequências potencialmente negativas são ignoradas ou atribuídas a algum outro fator. Grupos contra o programa, por outro lado, fazem exatamente o oposto. O debate público nos efeitos da intervenção tende a ser parcial, ideológico, egoísta e determinado por estratégias de *negação de culpa* e *reivindicação de crédito*. A conhecida falácia *post hoc, ergo propter hoc* (depois disso, portanto por causa disso) é conscientemente usada como munição política no debate público sobre os impactos do programa. A aferição de impacto é vista como um possível antídoto para essa tendência esmagadora de politizar os fatos.

² Nachmias (1979, p.7) cita H. M. Blalock: "Se X é uma causa de Y, temos em mente que uma mudança em X produz uma mudança em Y não meramente que uma mudança em X é seguida por ou associada a uma mudança em Y" (grifos dos autores)

Servidores públicos eleitos, gestores e avaliadores não deveriam tirar conclusões precipitadas, assim como o garoto do interior desta parábola. Dois meninos do interior viajaram para a cidade grande pela primeira vez em suas vidas. Antes de embarcarem no trem, eles compraram duas bananas, uma fruta que eles nunca haviam visto. Depois de um tempo, um deles começou a comer a banana e, enquanto ele comia, o trem passou por um túnel e o compartimento ficou completamente escuro. Horrorizado, o menino que comia a banana exclamou para o seu amigo: “não coma a banana, você vai ficar cego!”.

13.1 Delineamentos para aferição de impacto

Não há solução satisfatória e amplamente reconhecida para o problema de causalidade nas ciências sociais de forma geral, e muito menos em avaliação. Salvo alguns grandes esforços, as melhores mentes entre os cientistas sociais do mundo se provaram incapazes de fornecer mais que tentativas ambiciosas, porém geniais, de resolver essa questão.

O problema é, claro, que estados contrafactuais do passado não podem ser conhecidos com certeza. Pesquisadores não podem observar acima de qualquer dúvida o que teria acontecido se Napoleão tivesse vencido a Batalha de Waterloo, ou se o Imperador Francisco José não tivesse assinado a declaração de guerra da monarquia dual austro-húngara contra a Sérvia em 1914, ou ainda se o Japão não tivesse atacado Pearl Harbor. Podemos lembrar como Bertrand Russell colocou numa passagem do *Principia Mathematica*: “A escolha racional depende da construção de duas séries causais, das quais somente uma pode vir a existir. Portanto todo o estadismo (*statemanship*), e toda conduta racional da vida, é baseada no frívolo jogo histórico em que discutimos como o mundo seria se o nariz de Cleópatra fosse uma polegada mais longo.”³ Isso quer dizer que devemos discutir através de analogias mais ou menos confiáveis para separar os efeitos programáticos dos não programáticos.

³ A observação de Russell volta aos Pensamentos de Pascal: “se o nariz de Cleópatra fosse mais curto, toda a face da Terra mudaria”. Também cf. a celebrada pergunta contrafactual do historiador econômico Robert W. Fogel: “qual teria sido a taxa de crescimento econômico americano se as ferrovias nunca tivessem existido?” A publicação de **Railroads and American Economic Growth: essays in econometric history** (Baltimore, 1964) marcou a chegada da cliometria ou da Nova História Econômica Quantitativa. Mais tarde, Fogel recebeu o prêmio Nobel em Economia.

Na avaliação, o mundo ideal deveria ser uma cópia física perfeita da situação da intervenção. No delineamento experimental clássico, o grupo de controle constituiria uma duplicata do grupo experimental. Como o delineamento experimental perfeito é quase inatingível no mundo real das políticas públicas, abordagens aproximadas devem ser usadas. Existem várias dessas aproximações disponíveis para o avaliador. Em vários livros sobre métodos de pesquisa – escritos por autoridades como Rossi e Freeman (1999, p. 257-269), Judd e Kenny (1981), Nachmias (1979) ou Mohr (1995) – sustenta-se que as abordagens podem ser ordenadas de acordo com sua capacidade de resolver a questão de impacto. Essa ordem presumida é mostrada na figura 13.2. Contudo, também foi adicionada uma abordagem própria, estudos de caso (avaliação do processo explanatório).

Nas seções seguintes serão analisados alguns “delineamentos de pesquisa causal” – modelos lógicos de provisão de evidência – que supostamente permitem fazer inferências causais válidas (NACHMIAS, 1979, p. 21). Enquanto os delineamentos de pesquisa causal são abordagens para a análise de efeitos em geral, aqui eles foram adaptados para seus usos na avaliação de intervenção pública.

Figura 13.2

Delineamentos de Pesquisa de Avaliação para Esclarecer Impactos

Experimentos com Controles Randomizados (ECRs):

Medidas do resultado antes-depois (ou apenas depois) entre destinatários em um grupo – criado através da randomização e com desconhecimento que uma intervenção existe –, são comparadas às medidas do resultado antes-depois ou apenas antes de um grupo equivalente – também criado através da randomização – e que não conhecimento que a intervenção existe (experimentos randomizados clássicos)

Experimentos com controle pareados: Medidas do resultado antes-depois (ou apenas depois) entre destinatários da intervenção são comparadas com medidas do resultado antes-depois (ou apenas depois) de um grupo teoricamente equivalente, criado de forma não-aleatória através de combinações controladas, e dos quais a intervenção é omitida (experimentos não-randomizados).

Fig. 13.2 Continua na página seguinte

Controles Genéricos: Marcações dos resultados antes-depois para todos os destinatários da intervenção permanente que foram expostos a ela - seja pelo avaliador ou não - são comparadas com marcações dos resultados antes-depois típicas ou medianas que ocorrem em uma população maior razoavelmente equivalente não coberta pela intervenção e à qual se discute se os destinatários pertencam, ou da qual sejam um caso.

Controles Reflexivos: Dados do resultado antes-depois para algumas dimensões em todos os destinatários que receberam ou recebem a intervenção permanente realizada são comparados com um palpite com propriedade (*educated guess*) acerca dos dados do resultado *depois*, mostrando o que teria ocorrido nas mesmas dimensões e entre os mesmos destinatários caso a intervenção não ocorresse, e estimados com base nos dados do resultado antes da intervenção (extrapolação).

Controles-Sombra: Dados do resultado antes-depois entre destinatários que receberam a intervenção permanente são comparados com os julgamentos de pessoas com *insights* privilegiados (*experts*, gerentes do programa, elenco, participantes) sobre quais resultados eles acreditam que teriam acontecido sem a intervenção.

Estudo de caso (rastreamento do processo, avaliação do processo explanatório):

Para descobrir a extensão da influência da intervenção sobre os *outcomes*, estudam-se a formação, o conteúdo e a implementação da intervenção, a resposta dos endereçados (beneficiados/a quem se dirigiu a intervenção), a organização da função de controle, as ações dos *principais* depois da adoção da intervenção, e o contexto da intervenção como um caso rico em seus entornos naturais para descobrir e estabelecer fatores e mecanismos generativos na intervenção, em sua implementação e em seu contexto que funcionaram como facilitadores ou obstáculos.

Nota: uma intervenção permanente é uma intervenção “real” em contraste com um teste provisório.

Existem grandes divergências entre os teóricos da avaliação no que concerne aos delineamentos para aferir os efeitos das intervenções públicas. Particularmente no final da década de 1960 e dos anos 2000 até o presente, quando o movimento a favor das políticas públicas baseadas em evidências ganhou força, a experimentação baseada em dois grupos aleatórios foi tida como o padrão ideal. (CAMPBELL; STANLEY 1966, RIECKEN; BORUCH 1974, FAIRWEATHER; TORNATZKY 1977, SHADISH Jr; COOK; LEVITON 1991). Mas esse ponto de vista foi atacado veementemente, e estudos de caso, métodos interpretativos e rastreamento qualitativo do processo são recomendados em seu lugar.

13.2 Experimentos com controles randomizados (ECR)

O Cadillac da avaliação de programas –
experimentação aleatória e controlada.

**Harry P. Hatry et alli., Practical Program
Evaluation
for State and Local Governments, 2ª ed., 1981.**

Experimentação clássica e randomizada é caracterizada por uma lógica cristalina, parcialmente capturada na figura 13.3.

Figura 13.3

Experimentação Randomizada com medidas antes e depois do programa

Delineamento	Atribuição	Observação pré-programa	Programa	Observação pós-programa	Diferença
Grupo Experimental (E)	R	OE _{pre} (1010)	X	OE _{post} (1080)	E = OE _{post} - OE _{pre}
Grupo de Controle (C)	R	OC _{pre} (1015)	—	OC _{post} (1160)	C = OC _{post} - OC _{pre} (1160 - 1015 = 145)

R = atribuição ou designação randomizada

O = observação da variável resultado (variável dependente, variável alvo)

X = programa (política, projeto, elemento, instrumento de política, atividade, serviço e outras intervenções)

— = programa conscientemente omitido

O produto líquido do programa é E - C (70 - 145 = - 75, i.e., 75 unidades economizadas), onde:

OE_{pre}, OC_{pre} = marcações pré-programa [O representa observação(ões), medida(s)]

OE_{post}, OC_{post} = marcações pós-programa

E, C = resultado bruto para grupo experimental e de controle

O delineamento experimental clássico exige um *grupo experimental* e um *grupo de controle*, que devem se equivaler em todos os aspectos concebíveis. A maneira mais segura de estabelecer dois grupos sem diferenças sistemáticas é selecioná-los ao acaso. Um conjunto determinado de unidades – indivíduos, famílias, municípios, concessionárias, et cetera – são distribuídos em dois grupos de acordo com algum método, garantindo que a probabilidade de entrar em um dos dois grupos é exatamente a mesma para cada unidade. Esse processo – randomização – é marcado com um “R” na figura 13.3.⁴ Uma vez que a distribuição das unidades entre os dois grupos está feita, os valores da variável alvo – também chamada de variável dependente – são medidos nos dois grupos. Isso é marcado por OE_{pre} e OC_{pre} na figura, onde “O” significa “observação”; E significa “grupo experimental”; C significa “Grupo de controle” (*control group*), e “pre” para um número a princípio ilimitado de observações. O grupo experimental então é exposto ao programa enquanto o grupo de controle não é. Isso é indicado por X e - na figura 13.3. Para descobrir se o programa teve algum efeito, os valores da variável dependente são novamente registrados em ambos os grupos. Essas medidas são marcadas com OE_{post} e OC_{post} na figura, onde “post” significa um número ilimitado de observações. Mudanças que ocorreram entre as medidas antes e depois são consideradas causadas pelo programa.

Um exemplo planejado esclarece a lógica. Presumindo que desejamos saber os impactos de uma inovadora campanha governamental de informação para conservação de energia em domicílios com mais de uma família. Nas políticas públicas, a experimentação randomizada implica uma abordagem em dois estágios para o processo decisório. Antes de uma ação permanente, o programa deve ser experimentado provisoriamente numa escala pequena, exclusivamente sob condições experimentais. Com base nos achados produzidos pela avaliação experimental dessas tentativas provisórias, os tomadores de decisão decidirão se o programa deve ser posto em prática de forma permanente ou não.

⁴ Randomização não deve ser confundida com a amostragem aleatória. Na randomização, o pesquisador escolhe um conjunto limitado de unidades e *aloca* cada uma destas em dois grupos de acordo com algum método, o que garante a aleatoriedade. Por outro lado, a amostragem aleatória concerne à *seleção da amostra populacional* de forma que a amostra reflete todas as propriedades importantes da população. Os métodos podem ser combinados. Primeiro, podemos retirar uma seleção aleatória de unidades de uma população alvo e posteriormente alocar cada unidade em grupos experimentais ou de controle numa base aleatória (ROSSI; FREEMAN, 1999, p. 284).

Portanto, o procedimento em dois passos gera a conclusão que a avaliação deve ser planejada com antecedência e ser permitida a influenciar o projeto e a execução de uma campanha piloto. Isso quer dizer que o programa deve ser manipulável pelo avaliador; ele deve ter a capacidade de ligá-lo e desligá-lo. Se a pré-condição da manipulabilidade do programa é cumprida, dois grupos são selecionados aleatoriamente da população alvo total. Medições realizadas anteriormente mostram que os grupos usaram 1010 e 1015 unidades de energia antes do começo da campanha (Linha de base, os números são colocados na figura 13.3). A campanha agora é administrada ao grupo experimental, mas não ao grupo de controle. Medições posteriores mostram que o uso de energia aumentou em ambos os grupos, mas aumentou 75 unidades a mais no grupo de controle. A conclusão é que a campanha de economia de energia reduziu o uso de energia em 75 unidades, visto que sem a campanha, o uso de energia provavelmente seria 1155 no grupo experimental ao invés de 1080. Esse efeito pode ser razão o suficiente para discutir a favor da expansão do programa em todos os sentidos.

O cômputo do tamanho do efeito é realizado de acordo com a fórmula E-C, onde E é a diferença entre as mensurações depois e antes no grupo experimental e C é a correspondente dessa diferença no grupo de controle⁵.

Como o resultado de referência contrafactual é esclarecido na experimentação? Isso é feito ao produzi-lo de fato sob condições artificiais. Ao estudar o que está acontecendo com o grupo de controle não submetido ao tratamento, o avaliador extrai o que teria ocorrido caso o programa não tivesse sido posto em prática.

Como o experimentador pode alegar que os resultados são fruto do programa? A diferença entre os dois grupos pode ser devida a outro fato que ocorreu durante a experimentação. Graças à alocação aleatória de unidades nos dois grupos, experimentadores podem relaxar e ficar seguros que nenhum fator externo confundiu os resultados. Eles têm bons motivos para crer que *cetera* são *paria*, ou seja, que os grupos são aproximadamente equivalentes *em todos os outros aspectos*.

⁵ A experimentação randomizada só pode produzir equidade de grupos dentro de limites estatísticos conhecidos, devido a imperfeições do procedimento de randomização. Veja Rossi e Freeman (1999, p. 241); também Vedung (1997, p. 172 et seq).

Nenhuma outra circunstância além do fato que somente o grupo experimental foi exposto ao programa pode ter causado as mudanças observadas (WEISS, 1972a, p. 60 et seq., RIECKEN; BORUCH, 1974, p. 44 et seq., CAMPBELL; STANLEY, 1966, SUCHMAN, 1967, WHOLEY et alli., 1970). A condição *ceteris paribus* foi cumprida⁶.

ECRs (Experimentos de Controle Randomizado) são frequentemente considerados, embora longe de serem, em geral, a melhor solução para o problema do impacto na avaliação. Contudo, eles impõem pesadas demandas ao programa governamental. As intervenções devem primeiro ser testadas como experiências em pequena escala consumidoras de tempo com grupos experimentais e de controles, sendo acompanhada de uma avaliação. E há outras dificuldades. A consciência humana, o nosso poder de autodirecionar e automonitorar, é uma delas.

13.3 Efeito Hawthorne

As pessoas podem mudar seu comportamento uma vez que elas se tornam conscientes de seu papel como cobaias de um experimento. A sua interpretação do fato de que estão sendo introduzidos a um instrumento de medida pré-tratamento pode afetar seus resultados pós-tratamento. Os sujeitos da experiência podem se esforçar para fazer jus às expectativas dos pesquisadores, ou deliberadamente se comportar de forma contrária ao esperado. Para os especialistas, esse efeito é conhecido como *efeito Hawthorne*⁷. Essa estranha etiqueta de identificação é derivada da Hawthorne Works, uma unidade fabril da Western Electric Company em Cicero, Illinois. Entre 1924 e 1932, uma série de experimentos controlados foi lançada nesta fábrica da Western Electric, centrando-se no impacto do grau de iluminação na produtividade dos trabalhadores. No começo, tudo correu como esperado: quanto mais luz, mais produtividade.

⁶ "Ceteris Paribus" é uma frase latina traduzida literalmente como: "com todas as coisas [sendo] iguais", e é frequentemente grafada em português como "todo o resto é constante".

⁷ Outros rótulos são: efeitos reativos; reatividade (STONE, 2012, p. 198, NACHMIAS, 1979, p. 25, WEBB et al, 1966, p.13), *efeito cobaia*, ou "confusão inadvertida" (COOK; CAMPBELL, 1979, p. 38). Um expoente clássico da ideia de eliminar a reatividade nas Ciências Sociais é Webb et al (1966) em seu livro com o esclarecedor subtítulo **Nonreactive Research in the Social Science** ("Pesquisa Não-Reativa nas Ciências Sociais"). Deborah Stone (2001, p. 200 et seq) discute a insensatez de tal iniciativa.

Mas os pesquisadores descobriram, para seu espanto, que quando os processos eram revertidos, a produtividade continuava a aumentar. Quanto menos luz, maior a produtividade. Mulheres imigrantes trabalhavam arduamente durante 48 horas por semana sem intervalos, em estações de trabalho cada vez mais escuras, e ainda assim aumentavam sua produtividade. Perplexos, os cientistas não conseguiram encontrar uma explicação razoável. Contudo, entrevistas realizadas alguns anos depois revelaram que os funcionários gostavam de trabalhar em bancadas de trabalho separadas, desenvolveram sua vida social, conseguiram novos líderes e um novo objetivo comum. O importante foi: as trabalhadoras tiveram a impressão que a liderança da fábrica se importava com elas e tinha interesse em seu bem-estar. Na terminologia explicativa de hoje em dia, os pesquisadores descobriram os *mecanismos* nas mentes das trabalhadoras, conectando uma variável como iluminação a outra variável como a produtividade. Não foram as mudanças na iluminação, mas sim a consciência das mulheres que elas recebiam a atenção da liderança porque elas participavam em um experimento, ou, particularmente, sua interpretação do que essa atenção significava que governou seus comportamentos (WATT BOOLSEN, 2006, p. 77 et seq.; ANDERSON et al, 1974, p. 195 et seq.). De forma interessante, esse *insight* fundamental sobre motivos não-observáveis na cabeça das pessoas não pode ser capturado através da experimentação randomizada, somente depois de entrevistar (basicamente conversar com) os participantes acerca de seus pensamentos e suas interpretações.

Caixa 13.1: Profecias autorrealizáveis

O Efeito Hawthorne está ligado a profecias autorrealizáveis. Uma profecia autorrealizável é uma definição falsa de uma situação que evoca um comportamento novo que torna verdadeira a concepção anteriormente falsa.

“Se os homens definem as situações como verdadeiras, elas são reais em suas consequências.” Robert Merton começa sua clássica exposição na profecia autorrealizável (1968, p. 476 et seq.) com essa citação de W. I. Thomas. Para ilustrar o teorema de Thomas, Merton conta a trágica história do florescente *Last National Bank* de C. Millingville. Começou na Quarta-feira Negra de 1932, ‘a última quarta-feira do *Last National Bank*, deve ser notado.’ O Banco era uma instituição bem gerenciada. Uma grande parte de seus recursos era líquida sem ser dissolvida. Mas um rumor de insolvência surgiu. Uma vez que um número suficiente de depositantes passou a acreditar nesses rumores, eles começaram a retirar freneticamente seu dinheiro, que logo acarretou a insolvência do banco. Uma vez que os depositantes definiram a situação como ruim, apesar de ela ser boa, a sua opinião da situação se tornou real em suas consequências.

Nos assuntos humanos, não só as “condições objetivas” importam, mas o significado atribuído a elas pela consciência humana também tem seu papel. Merton (1968, p. 477) discute:

Definições públicas de uma situação (profecias ou previsões) se tornam parte integral da situação e, portanto, afetam os desenvolvimentos subsequentes. Isso é peculiar para os assuntos humanos. Não é encontrado no mundo na natureza, ainda intocado pelas mãos humanas. Previsões do retorno do Cometa Halley não influenciam a sua órbita. Mas a suposta insolvência do banco de Millingville de fato afetou o resultado. A profecia do colapso levou a seu próprio cumprimento.

A falência do *Last National Bank* é um exemplo de profecia autorrealizável. Apesar da liquidez comparativa dos bens do banco, um rumor de insolvência, tido como verdadeiro por um número suficiente de depositantes, levou à insolvência real do banco.

Longe de se render ao Efeito Hawthorne, defensores da experimentação randomizada encontraram jeitos de neutralizá-lo. O sistema duplo cego (*double blindness*) é um deles. Um experimento é duplamente cego se informações que podem levar a uma parcialidade dos resultados são escondidas tanto do experimentador quanto dos participantes dos dois grupos de teste. Por exemplo, os participantes não deveriam saber se eles pertencem ao grupo experimental ou ao de controle; tampouco o experimentador deve saber quem pertence a qual grupo. Para garantir a “cegueira” entre os participantes, os membros do grupo de controle são sujeitos a um *placebo* (lat. “eu devo agradar”), um tratamento falso, que é passado como idêntico ao verdadeiro administrado aos membros do grupo experimental, enquanto ele é na verdade idêntico somente em certos aspectos superficiais e não potentes. Um placebo dá a aura de um tratamento real, mas não a substância. O uso de um placebo é completamente enganoso.

Apesar de ser um recurso frequentemente utilizado em laboratórios médicos, o sistema duplo cego com placebos não pode ser aplicada regularmente em grande escala nas políticas públicas. Primeiro, os cidadãos provavelmente não aceitariam ser enganados dessa forma pelos seus governos; segundo, eles contradizem a norma de abertura social e transparência em uma democracia.

O Efeito Hawthorne é uma deficiência em potencial de toda experimentação social, mesmo em condições de laboratório. Além disso, também há outros argumentos bastante fortes contra a experimentação randomizada nos testes-piloto (*try outs*) do setor público – motivos que serão discutidos em uma seção separada mais à frente.

Confrontados com essas dificuldades, defensores da Experimentação com Controle Randomizado (ECR) relutantemente dão um passo para trás e recomendam controles combinados como uma alternativa aceitável. Apesar de produzirem achados mais fracos nos efeitos da intervenção, controles combinados são mais adaptados à tomada de decisões do setor público.

13.4 Experimentos com Controles Pareados

Mais e mais eu cheguei à conclusão que o centro do método científico não é a experimentação em si, mas sim a estratégia conotada pela frase “hipóteses rivais plausíveis”.

Donald T. Campbell, Preface, in Yin: Case Study Method.

Os experimentos com controles pareados ou combinados também são baseados na abordagem em dois passos para a criação de políticas públicas. Primeiro, as intervenções públicas preliminares são introduzidas em testes de pequena escala (projetos-piloto) e cuidadosamente avaliadas através do uso de dois grupos combinados. Segundo, os achados produzidos pela avaliação devem prover orientação razoável quando os tomadores de decisão determinarem se as intervenções (modificadas) serão implementadas numa escala compreensiva ou não. Nos controles pareados, contudo, nenhum dos dois grupos no teste piloto são criados de forma aleatória. Ao invés disso, o avaliador seleciona um grupo de controle intacto e que lembra de perto o grupo experimental com dimensões premeditadas. Idealmente, o outro grupo deveria ser igual ao grupo experimental em todos os aspectos considerados relevantes. A analogia perfeita para a equação seria um par de gêmeos *monozigóticos*, criados pelos mesmos pais sob circunstâncias sociais e naturais idênticas.

O termo profissional herdado para esse processo é *pareamento* (*matching*). Na Índia e na Coreia, por exemplo, os pais devotam uma enorme quantidade de energia para o casamento arranjado. Eles tentam arranjar seu filho ou filha para um parceiro apropriado. Experimentos com controles não-aleatórios são afirmados a partir do pareamento.

Contrário ao que é sustentado por vários especialistas, a manipulabilidade da variável de entrada – a intervenção, o programa – não constitui a fronteira entre experimentos randomizados e combinados (CAPORASO; ROOS, 1973, p. 39). Tanto nos experimentos randomizados quanto nos combinados, o pesquisador pode decidir quanto ao *timing* da intervenção. Em vez disso, a diferença decisiva envolve como a equação entre os dois grupos é atingida. Nos experimentos de controle randomizados, a equivalência é estabelecida através da randomização; em experimentos combinados, através de pareamento inteligente.

O outro gêmeo idêntico criado sob condições sociais equivalentes provavelmente nunca será encontrado pelo avaliador.

Ele terá que se contentar com pares, que são apenas aproximadamente equivalentes aos grupos experimentais em certos aspectos relevantes. E quais são esses aspectos depende de considerações teóricas relevantes à questão feita pelo avaliador em sua avaliação.

Se o problema é detectar até onde uma campanha governamental de comunicação produzirá de fato uma redução no consumo de energia nos domicílios, o pareamento deve proceder com referência a *todas* as contingências que possivelmente afetarão o uso de energia por domicílio. O avaliador terá de formar uma opinião com respeito a quais fatores causais podem influenciar na utilização de energia em um domicílio. O pareamento deve ser guiado por uma boa *teoria da situação* (*teoria do campo da intervenção, teoria de problema*, indicando os determinantes das mudanças no fenômeno a ser observado (ROSSI; FREEMAN, 1999, p. 315 et seq.). O pareamento é informado pela teoria.

Nesse exemplo, baseado em avaliações anteriores, pesquisas e em experiências empíricas, temos que moldar uma opinião bem-fundamentada do que faz alguns domicílios mais dispostos a conservar energia que outros. Alguns fatores potencialmente influentes estão na figura 13.4

Figura 13.4

Teoria da Situação do Campo de Conservação de Energia por Domicílio

-
- 1 Preço da energia**
 - 2 Tamanho do domicílio**
 - 3 Idade do domicílio**
 - 4 Tipo de domicílio: simples, família, condomínio, aluguel**
 - 5 Medidores em conta coletiva ou conta individual**
 - 6 Equipamentos domésticos**
 - 7 Número de pessoas em casa durante o dia**
 - 8 Renda**
 - 9 Idade dos membros do domicílio**
 - 10 Atitude quanto a preservação de energia e proteção do ambiente**
 - 11 Existência de zelotes**
-

No pareamento, os avaliadores tem que ter certeza que os grupos são tão equivalentes quanto possível nesses aspectos em particular. Ao garantir que essas circunstâncias operam *com forças iguais* em ambos os grupos, o experimentador pode permitir-se desconsiderá-las. Por meio disso, alguns fatores explanatórios específicos que se constituem em rivais plausíveis para o programa podem ser deixados de lado. (CAMPBELL; STANLEY, 1973, p. 204 et seq.; WEISS, 1972b, p. 69 et seq.).

Por isso, um pareamento devidamente executado não é um assunto fácil ou rápido. Ele pressupõe conhecimento teórico e habilidades técnicas. Seriam esses pareamentos impraticáveis e complicados demais para serem executados num cenário de políticas públicas e administração pública?

13.5 Controles Pareados ou Concordantes: O Delineamento de Séries Temporais de Controle

Mais poderoso entre os delineamentos combinados por pareamento (concordantes) com referência à causalidade é o Delineamento de Séries Temporais de Controle. Outra abordagem, mais simples, é o Delineamento com Grupo de Comparação Pré-teste/Pós-teste, é uma versão amputada do Delineamento de Séries Temporais de Controle (Figura 13.5). O enfoque agora será no Delineamento de Séries Temporais de Controle.

Figura 13.5 - Dois Experimentos Pareados ou Concordantes

Delineamento	Atribuição	Observação pré-programa			Programa	Observação pós-programa		
<i>Delineamento de série de controle</i>								
Grupo de investigação	M	O ₁	O ₂	O ₃	X	O ₄	O ₅	O ₆
Grupo de comparação	M	O ₁	O ₂	O ₃	—	O ₄	O ₅	O ₆
<i>Delineamento de grupo de comparação de pré / pós-teste</i>								
Grupo de investigação	M			O ₁	X		O ₂	
Grupo de comparação	M			O ₁	—		O ₂	

M = atribuição ou designação por meio do pareamento
O = observação (mensuração) da variável resultado (variável dependente, variável alvo)
X = programa (política, projeto, elemento, instrumento de política, atividade, serviço e outras intervenções)
— = programa conscientemente omitido

Fonte: Adaptado de uma figura em Riecken e Boruch (1974, p. 98).

O Delineamento de Séries Temporais de Controle envolve a coleta de dados da variável alvo em dois grupos, feitos para se parecerem um ao outro no maior número de dimensões possível. Os pontos dos dados deveriam formar uma série temporal e englobar tanto a era pré-programa quanto a pós-programa para ambos os grupos. Aproximadamente no meio do período de tempo, um grupo foi naturalmente introduzido ao programa, enquanto o outro não. Diferenças nas marcações após o programa podem ser atribuídas ao mesmo programa. Essa conclusão baseia-se no fato que os dois grupos são aproximadamente equivalentes no que tange todos os fatores que podem influenciar a variável do resultado – uso de energia – e estão, portanto, expostos à *mesma* influência externa desses fatores. Mas existe uma diferença crucial entre esses dois grupos: os grupos experimentais foram introduzidos a um programa enquanto os de controles não. Isso deve explicar diferenças eventuais nas marcações pós-programa.

Experimentadores prontamente concedem que eles não podem ter certeza absoluta se a explicação do programa se sustenta ou não. Apesar do mesmo também valer para os controles randomizados, a insegurança nos controles combinados é maior. No vocabulário técnico da avaliação, eles admitiriam que a explicação do programa é posta em risco por fatores explanatórios que competem entre si. Por isso, Campbell e Stanley (1966, p. 5 et seq.) cunharam a famosa frase “ameaças à validade interna”.

13.6 Ameaças à Validade Interna

Antes de aceitar a explicação do programa como plausível, o avaliador que usa controles combinados deve considerar sistematicamente como várias ameaças à validade interna podem ter influenciado o resultado. Ele deve, talvez, acumular dados esclarecendo o tamanho e o efeito das ameaças e fazer uso de todo seu conhecimento teórico, senso crítico e poderes de inovação⁸.

⁸ Campbell e Stanley (1963) e Cook e Campbell (1979, p. 37 et seq) fornecem relatos sistemáticos, porém técnicos, de ameaças. A **Encyclopedia of Evaluation** (MATHISON, 2005, “Validity”) lista 13 ameaças: história, maturação, testes, instrumentação, regressão estatística, seleção, mortalidade, equalização compensatória de tratamentos, difusão ou imitação de tratamentos, interação da seleção e tratamento experimental, e esforços reativos de arranjos experimentais. Bryman (2004, pp. 36-37) lista história, maturação, testes, instrumentação, mortalidade e ambiguidade relativos à direção causal. Para uma apresentação mais simples, consultar Herman (1987, parte 3).

Caixa 13.2 – Delineamento de Série Temporal de Controle Fatores a Considerar e Ameaças para Evitar

Existem algumas ameaças às quais o Delineamento de Série Temporal de Controle deveria ser imune, supondo que medidas antes e depois têm sido coletadas por um período de tempo suficientemente longo.

A mudança não pode ser atribuída à *instabilidade aleatória*. Uma longa série de tempo inexoravelmente revela variações acidentais nas variáveis alvo. Segundo, a mudança não pode ser parte de uma *tendência em longo prazo*. A série temporal estendida dos dados para ambos os grupos nos diria se esse fosse o caso. E terceiro, a mudança não pode ser parte de um padrão se repetindo. Variações sazonais são bons exemplos. Quando as séries de dados são suficientemente longas para cada grupo, podemos facilmente ver se as mudanças nos resultados estão envolvidas em algum padrão periódico.

Por contraste, a mudança pode depender de *procedimentos de registro de dados modificados*. Mudanças na montagem dos dados e manutenção de registro geram idiosincrasias nas medidas. Caso um ajuste de fato seja realizado, é possível que os novos dados se tornem comparáveis aos antigos.

Mais importante, todavia: as mudanças podem resultar de *fatores substantivos* além do programa. Algum fator causativo importante deve ter sido ignorado no pareamento. Esforços prolongados à parte, a avaliadora pode ter produzido um par imperfeito. Alguma contingência substantiva fora do controle da avaliadora pode ter impactado suas marcações das variáveis alvo.

Em outras palavras, as soluções para o problema da causalidade fornecidas pelo Delineamento de Séries Temporais de Controle nunca são melhores que a teoria da situação de fato empregada no pareamento. Contudo, este pareamento nunca é exaustivo. Para aferir ameaças substantivas não eliminadas pelo pareamento, o avaliador tem que coletar dados em cada um deles separadamente e sujeitar esses dados a análises estatísticas ou raciocínio perspicaz para decidir se esses fatores foram influentes ou não. Tal análise de dados suplementares é complicada e consome tempo. Levando em conta esse cenário, a experimentação combinada é realmente factível na avaliação do setor público?⁹

13.7 Programas Públicos Deveriam Ser Experimentação Social?

Antes que decisões precisas sobre a efetividade de algum programa social possam ser feitas, é bastante óbvio que deve haver uma avaliação completa e científica dele. É ... importante que a avaliação seja muito saudável de um ponto de vista científico. Isso, em essência, quer dizer que a avaliação deve ser de fato uma avaliação *experimental*.

G W Fairweather & L G Tornatzky, *Experimental Methods for Social Policy Research*, (1977).

O preço que você paga pela precisão é a inabilidade de lidar com perguntas do mundo real.

Douglass North (*Prêmio Nobel de Economia*, 1974)

⁹ O “modelo de coortes” ou “modelo de apenas grupos não equivalente no pós-teste” se assemelha aos controles pareados porque ambos são modelos de dois grupos, e em ambos, o tratamento (a intervenção) é manipulado pelo avaliador (experimentador). Uma diferença é que as unidades incluídas no experimento decidiram pessoalmente se elas querem pertencer ao grupo experimental ou de controle. Uma amostra de unidades voluntárias é dividida em dois grupos ao permiti-los ser expostos voluntariamente ou não serem expostos à intervenção. Como a seleção se baseia na escolha voluntária de participar, nem controles randomizados nem pareados estão envolvidos. A seguir, o avaliador segue os membros individuais dos dois grupos no futuro. O avaliador não segue a categoria em si, mas os indivíduos que se enquadram na categoria em si. Apenas medidas após a intervenção são tomadas, sem medidas pré-intervenção.

A experimentação aleatória controlada (ECR) foi adotada como o Cadillac da avaliação do setor público. Particularmente de 1965 até o início dos anos 70 e novamente nas décadas de 2000 e 2010, foi e é tido como a melhor marca que a metodologia das ciências sociais pode contribuir para a avaliação. A força presumida do projeto RCT como um método de intervenção do setor público foi formulada de forma clara e sucinta por Riecken e Boruch há mais de quarenta anos atrás:

Nós... enfatizamos o papel da informação obtida experimentalmente na formação de uma política social porque tal informação é bastante útil no aprendizado das relações de causa entre elementos do programa e *outcomes*. Se um efeito pode ser demonstrado num grupo de unidades (pessoas, lugares, ou instituições) escolhido aleatoriamente e sujeito a um tratamento específico enquanto um grupo similar, que não é tratado, não mostra qualquer efeito, pode-se ficar razoavelmente confiante que o tratamento produziu tal efeito. Tal confiança não pode tão prontamente repousar numa evidência não-experimental, apesar de métodos sofisticados de análises podem ser utilizados para reduzir a ambiguidade da inferência causal. A superioridade do método experimental reside no fato de que num experimento verdadeiro as diferenças entre um grupo tratado (experimental) e um grupo não-tratado (controle) pode ser atribuído completamente ao efeito do tratamento somado a um componente acidentalmente errado (acaso) que pode ser estimado com precisão e que será igualmente distribuído pelos grupos de controle e experimental. Além disso, todos os outros fatores que aumentam ou diminuem o *outcome* variável ocorre igualmente em ambos os grupos, experimental e de controle. (RIECKEN; BORUCH 1974, p. 5)

Contudo, a experimentação randomizada é pesadamente criticada. Efeitos Hawthorne podem danificá-la quase tão severamente quanto eles danificam outros delineamentos. Mais sério que essa

objeção geral, contudo, são as acusações de que a experimentação é inapropriada como um *delineamento para avaliação de políticas públicas*. Experimentos levantam demandas tão pesadas na estrutura das intervenções públicas de maneira a torná-las quase impossíveis de serem usadas. As pré-condições, que o grupo alvo seja inicialmente dividido entre grupo experimental e grupo de controle, e que o primeiro seja exposto a uma intervenção provisória em um teste piloto sem que o grupo de controle seja informado, antes que a intervenção permanente seja efetuada, raramente são satisfeitas no mundo real da política pública, onde programas de política pública são, por via de regra, permanentemente efetuados em escala abrangente. Não haverá grupo de controle de quem os programas serão acobertados. Isso está conectado a uma característica importante da criação de políticas públicas. Atores administrativos e políticos não podem adotar intervenções em um isolamento esplêndido quanto ao que está acontecendo em volta deles. Servidores públicos eleitos tem que levar em conta as opiniões e ações de outros jogadores, como a mídia, os eleitores, atores econômicos no mercado, e potências estrangeiras. A necessidade de ação rápida torna a experimentação planejada rígida e consumidora de tempo. Os requisitos que as intervenções sejam experimentadas de forma provisória com grupos aleatórios ou pareados (concordantes) antes de sua inauguração completa não podem ser cumpridos no mundo real da programação do setor público.

Desse dilema existem dois, e apenas dois, estratégias. Ou o processo decisório do setor público deve ser adaptado à pesquisa de avaliação radical, ou a pesquisa deve ser adaptada ao processo decisório do setor público.

Alguns teóricos da avaliação profundamente interessados escolheram a primeira opção. Servidores públicos devem se aprumar e agir de uma forma mais racional. Estes teóricos não estão preocupados com melhoras marginais dos programas atuais, mas com conjecturas audaciosas e grandes reformas inovadoras em uma sociedade utópica do futuro. Esses programas inovadores devem ser ajustados às demandas dos avaliadores pelos melhores delineamentos científicos disponíveis. Decisões sobre a implantação de intervenções completas devem ser precedidas por testes-pilotos, delineados como experimentos combinados ou randomizados. E igualmente importante,

os tomadores de decisão devem prestar atenção ao conhecimento avaliativo na estruturação de programas permanentes. A avaliação rígida deve receber um papel decisivo na formação de intervenções públicas permanentes.

Mais do que ninguém, talvez, Donald T. Campbell, o ilustre metodologista da avaliação americano, expôs o sonho racionalista de programação do setor público como uma experimentação qualificada e contínua sob a direção das ciências sociais. Em seu famoso artigo “Reformas como Experimentos” (CAMPBELL, 1969, p. 409), ele sustentou, entre outras coisas, o seguinte:

Os Estados Unidos e outras nações modernas deveriam estar prontos para uma abordagem experimental com respeito à reforma social, uma abordagem para a qual se testa novos programas desenhados para curar problemas sociais específicos, nos quais aprendemos se esses programas são efetivos ou não, e os quais serão retidos, imitados, modificados ou descartados com base na efetividade aparente.¹⁰

Campbell promoveu constantemente o *policy-making* experimental no setor público e esta posição pertence a uma escola de pensamento semi-tecnocrática, de acordo com a qual o *policy-making* deveria ser mais científico, senão modelado nas ciências. Kurt Lewin, o eminente psicólogo social, mantinha uma visão semelhantemente otimista do papel dos experimentos para a melhora social. Um famoso expoente radical dessa linha de pensamento era o físico inglês John Bernal, que na década de 1930 apresentou grandes sonhos visionários de políticas públicas conduzidas de acordo com princípios científicos, e com cientistas eminentes ocupando posições de destaque no governo. Campbell nunca chegou tão longe; em sua tentativa de utopia, cientistas sociais são apenas servos metodológicos dos tomadores de decisão, não conselheiros nem governantes.

¹⁰ Também Campbell (1970, 1982) e Campbell e Stanley (1966) e Cook e Campbell (1979). Essa tese também foi defendida por Suchman (1967), Rivlin (1971), Riecken e Boruch (1974), James S. Coleman (1964) e Hans L. Zetterberg (1964). A filosofia de Campbell é analisada em Shadish Jr., Cook e Leviton (1991).

Deveria a formulação de políticas públicas ser empreendida em dois passos? Deveriam intervenções governamentais audaciosas ser experimentadas de forma provisória em experimentos sociais realistas e sua aplicação permanente depender dos resultados desses experimentos? À primeira vista, essa visão em dois estados de um sistema político racional parece sã. Por que não se deveria, usando a melhor metodologia disponível, descobrir se os programas propostos realmente atingirão seus objetivos antes de serem inaugurados em escala abrangente?

Contudo, a experimentação possui algumas fraquezas como um instrumento de planejamento público, um fato prontamente admitido pelos próprios experimentalistas. Experimentos consomem tempo, produzem conhecimento restrito e são fracos na validade externa.

Experimentos de laboratório e testes de campo consomem tempo. Essa objeção é exposta por Cook e Campbell (1979, p. 244) da seguinte maneira:

Geralmente uma quantidade considerável de tempo se passa entre o planejamento de um experimento randomizado e a obtenção de seus resultados. As perguntas de pesquisa mais importantes tem que ser desenvolvidas e validadas; trabalho piloto (ou pesquisa formativa) deve ser conduzido; as medidas pré-teste devem ser realizadas; o resto dos dados deve ser coletado, limpo e ordenado; a análise deve ser conduzida; e os primeiros rascunhos dos relatórios devem ser cuidadosamente examinados. Portanto, não deveria ser surpreendente que tipicamente anos se passem entre a concepção de um experimento e a disponibilidade dos resultados. Quando as decisões devem ser feitas rapidamente, experimentos randomizados podem não ser adequados.

Os políticos podem realmente esperar anos antes de agir? Ocasionalmente sim. Tomemos o caso de encontrar um repositório final seguro para lixo nuclear radioativo. Pesquisas nessa questão

acontecem há décadas e podem continuar por muitos anos. O problema é tão sério e a perspectiva de tempo é tão longa que o período de planejamento político pode ser estendido.

Na maior parte das ocasiões, contudo, a resposta é claramente “não”. É óbvio que corpos políticos devem reagir prontamente a desastres naturais imprevisíveis como terremotos, enchentes, secas, tornados, deslizamentos de terra, avalanches e tempestades de neve. Mas o Estado também deve agir velozmente em resposta a expectativas alimentadas ou medidas tomadas por outros agentes. Por exemplo, os partidos incumbentes no governo frequentemente pensam que tem que tomar ação imediata porque eles prometeram ao eleitorado que o fariam. Frequentemente, o Estado deve adotar medidas rápidas devido a movimentos repentinos do capital entre fronteiras. O exemplo padrão são as medidas para reduzir a fuga de capitais quando se começa a acreditar em rumores de desvalorização de moeda. Experimentalistas radicais sugerem visões não-realistas e até ingênuas da natureza firme e de longo alcance do *policy-making* público (processo de formulação de políticas públicas).

Segundo, experimentos produzem *conhecimento restrito* em vários sentidos. Na experimentação, é difícil discernir os efeitos de mais de um programa, porque a administração de um programa extra, porém extremamente parecido, envolve a complicada utilização de outro grupo selecionado aleatoriamente. Pela mesma razão prática, processos entre a adoção do programa e os resultados não são esboçados nos experimentos. Os mecanismos mais refinados das respostas formais do implementador, intermediador e beneficiário são tratados como caixas pretas. Isso poupa tempo e trabalho preciosos do avaliador¹¹. Para os servidores públicos, contudo, é uma desvantagem séria, já que a informação contida nos mecanismos que movem as considerações e ações tanto do implementador como do beneficiário são de interesse substancial para eles. A experimentação, pelo que parece, não pode substituir a sabedoria rica e variada adquirida através de outras fontes.

¹¹ Modelos de experimentos fatoriais que envolvem mais de uma variável de programa provavelmente são complexos demais para serem manuseados em políticas públicas. Veja Nachmias (1979, p. 32 et seq).

Uma última injunção concerne à *dependência do contexto*. Em cenários de campo complexos, aferidores de impacto não podem fazer a suposição geral que os achados de um experimento artificial se manterão verdadeiros no campo. Processos governamentais reais são provavelmente mais diferentes das situações experimentais no laboratório que qualquer outra situação social à qual possamos querer aplicar os achados dos experimentos. “Mesmo na agricultura, a transição do experimento controlado na estação de pesquisa para a prática média na fazenda é sempre problemática. Mesmo que o fertilizante A seja superior ao B em experimentos controlados, ele pode não ser superior no tipo de solo especial encontrado, digamos, no Sudeste da Geórgia.” (COOK; CAMPBELL, 1979, p. 7).

No jargão da avaliação, experimentos randomizados e combinados são fracos na *validade externa (generalização para outras situações)*, CAMPBELL; STANLEY, 1966, p. 5 et seq.). Em situações experimentais, os resultados são, a princípio, independentes do contexto. Mas no mundo real, o contexto possui um papel decisivo. Transferir resultados experimentais para grandes cenários de campo é extremamente problemático.

A Experimentação é certamente um *projeto* excelente se aplicado nas ciências. Também pode ter seu valor em algumas ciências sociais, particularmente a psicologia e a sociologia de grupos pequenos. No governo e na administração pública, a situação é diferente. De fato, existe forte resistência contra a utilização da programação experimental em políticas públicas.

Todavia, experimentos, testes de campo provisórios e programas piloto são estabelecidos (BERK et al, 1985, HELLSTERN; WOLLMAN, 1983). Nos Estados Unidos, vários dos famosos programas da Grande Sociedade de Lyndon B. Johnson foram inicialmente lançados como testes. Alguns deles também foram avaliados através da experimentação controlada. De acordo com Charles Murray (1984, p. 149), a inauguração do Programa de Imposto de Renda Negativo – um programa que fornecia pagamentos a pessoas cuja renda caía abaixo de certo piso:

“tomou a forma do experimento de ciências sociais † mais ambicioso da história. Nenhum outro sequer chega perto da sua combinação de tamanho, gastos, abrangência no tempo e detalhe de análise... Começou em 1968, utilizou em seu fim 8700 pessoas como sujeitos, e durou dez anos... Resultou em um corpo de literatura que, na década de 1980, incluía mais de cem títulos publicados e incontáveis relatórios não publicados. Seus custos foram de muitos milhões.”

Ambições à parte, a contribuição da experimentação para o programa foi ou ínfima ou nula.

Outros países além dos Estados Unidos usam experimentação em uma escala considerável. Isso se aplica particularmente à Dinamarca e à Noruega, por exemplo. Também há exemplos históricos. Na monarquia austro-húngara, reformas eram introduzidas como testes na Galícia antes de ser realizadas de forma integral no resto do império.

Qualquer abordagem experimental ou semelhante a um experimento é frequentemente considerada típica do *policy-making* sueco. Testes de campo, demonstrações e protótipos são lugar-comum na vida política sueca. As grandes reformas escolares da década de 1950, por exemplo, foram experimentadas em larga escala em um número de comunidades antes de serem aplicadas no território nacional.

Contudo nem mesmo o uso amplo de testes e protótipos sueco pode ser levado como evidência que essa sociedade é um laboratório social qualificado. Enquanto frequentemente se referem a eles como “experimentos”, a maior parte dos testes de campo não são delineados nem como experimentos randomizados, nem como experimentos completamente concordantes. Na maior parte das ocasiões, talvez, o motivo operativo não é experimentar e avaliar de uma forma imparcial; às vezes são gestos simbólicos projetados para indicar que o governo possui ciência dos problemas para coletar votos e suporte da opinião.

Ocasionalmente, o governo quer desbravar uma trilha para a adoção em escala completa da reforma, independente de como ela funcione. Tendo previamente decidido instigar uma reforma completa, os *decision-makers* não tem a influência política para realizá-la de forma imediata e completa. Ao experimentá-la de forma supostamente

provisória, a questão vai amadurecer e a resistência será amortecida. Em realidade, os supostos testes de campo experimentais envolvem pouca coisa além da construção de consenso político sobre a reforma. Demonstrações podem se tornar uma espécie de estratégia de implementação (HADENIUS, 1990, p. 80, p. 85 et seq., p. 117 et seq.; PREMFOR, 1989).

Além disso, testes de campo preliminares são usados para legitimar as reformas. Para esse propósito, os testes de campos que coletam dados sobre a implementação e impacto não são necessariamente importantes, mas sim o fato de que as rodadas piloto foram empreendidas. Testes de campo facilmente tomam um ar de ritual racional e manobras táticas. Não se pode nunca deixar passar a possibilidade que a avaliação pode ser usada para legitimar algo previamente decidido ou estabelecido.

É importante distinguir o caso para experimentação nas ciências naturais, nas ciências sociais básicas e em políticas públicas e avaliação de programas. Até onde se sabe, em alguns campos das ciências, a experimentação é amplamente aclamada como o projeto que deve ser aplicado. O mesmo se mantém verdade, mas numa extensão consideravelmente menor, com algumas ciências sociais básicas como a psicologia e a sociologia de grupos pequenos. Na administração pública e governamental, a situação é diferente. Apesar de um cenário bastante restrito, a experimentação tem seu papel, porém ele é provavelmente apenas marginal.

Uma possível abordagem para a experimentação seria criar alguma divisão do trabalho entre pesquisa acadêmica e políticas públicas/administração pública. As universidades deveriam realizar mais experimentação do que vêm fazendo no presente para descobrir "o que funciona". Os achados, então, deveriam ser disseminados aos profissionais do setor público para ajustes e possivelmente implementar mudanças. Isso parece ser a saída escolhida pelo movimento a favor do *policymaking* e administração pública baseados em *evidências*.

Dentro do setor público, uma distinção mais nítida deve ser feita entre política de alto nível e intervenções administrativas. Seria mais fácil conduzir a experimentação com propostas administrativas de mudança de baixo nível do que com reformas políticas de alto nível.

Além disso, pesquisadores da avaliação devem aceitar que a formação de políticas públicas não pode nunca ser uma ciência sonhada

por alguns experimentalistas radicais. Os métodos de avaliação devem ser ajustados ao jogo político democrático, e não o oposto. Isso não implica que a pesquisa da avaliação é fútil. Deve haver uma preocupação com a avaliabilidade no estágio de planejamento de uma intervenção pública. Contudo, não se pode permitir que essa preocupação se torne primordial – um caso onde a avaliação é mais importante que o projeto, empurrando a formulação democrática das políticas públicas para baixo. Se forem politicamente aceitáveis, testes preliminares podem e devem ser arranjados. Eles também devem ser avaliados, ora pela experimentação radical, ora pela participação dos *stakeholders* e de métodos dialógicos. Novas intervenções inovadoras e reconsiderações fundamentais das intervenções existentes ocorrem de vez em quando na vida pública, e nessas situações, a experimentação pode ter seu papel além de outros delineamentos de avaliação. Por outro lado, os avaliadores devem aceitar que mesmo nos testes preliminares de campo de inovação, a experimentação frequentemente se encontra fora de questão.

Em primeiro lugar, e mais importante, é o fato de que os avaliadores devem aceitar que a maior parte das intervenções governamentais já está ocorrendo e não pode estar sujeita a testes de campo. Mudanças ocorrem, mas elas são incrementais. Nesses casos, os avaliadores devem trabalhar em delineamentos que são mais fracos ao estabelecer a equação entre os dois grupos, mas mais facilmente aplicados à prática política existente. Deve-se colocar menos ênfase na experimentação de dois grupos e mais nos controles reflexivos, em técnicas transeccionais e sombra e, acima de tudo, avaliações de processo de um tipo qualitativo, de estudo de caso, e avaliações que buscam mecanismos explicativos (*qualitative, case-study, mechanism-seeking type*). No próximo capítulo serão examinados esses modelos avaliativos.

CONTROLES GENÉRICOS, REFLEXIVOS E CONTROLES-SOMBRA

Quando a intervenção é adotada e a seleção em curso de dois grupos razoavelmente equivalentes para uso em experimentação é problemática, os avaliadores têm de empregar os chamados projetos mais fracos, mais adaptados às realidades políticas.

Os projetos recomendados como alternativas são *naturalistas*¹. Concentrando-se em intervenções que vão se desenrolando na vida real, não é feita nenhuma tentativa de adequá-los às exigências de experimentos aleatórios ou combinados. Políticas, programas, instrumentos de política, projetos e elementos são examinados em seus contextos naturais, permanentes, e não em configurações de teste provisórias. O espírito de laboratório desaparece porque experimentos e controles não são nem artificialmente criados (como em experimentos aleatórios), nem combinados (como em quase experimentos).

Para Rossi, Freeman e Lipsey (1999, p. 343 et seq.) as alternativas naturalistas são duas: controles reflexivos e controles-sombra. Eles serão complementados aqui com controles genéricos, um projeto que Rossi, Freeman e Lipsey tratam como quase experimental. Como quarta concepção naturalista serão acrescentados estudos de caso, caracterizados por rastreamento de processos, e, na linguagem avaliativa, rotulado de avaliação de processo. Os estudos de caso serão expostos no capítulo 15 à parte.

14.1 Controles Genéricos

Um pré-requisito para o projeto de controles genéricos é que nós temos pelo menos um, de preferência várias medidas consecutivas, mostrando marcações pré-intervenção sobre a variável resultante (linha de base) e pelo menos um, de preferência com medidas variadas, que mostra marcações pós-intervenção.

¹ Na verdade, até mesmo um tipo de experimento é um projeto quase naturalista, na medida em que ele estuda o que aconteceu como consequência da intervenção permanente e compara esse *ex-post-fato* a um grupo de controle pareado.

Os resultados reais deste grupo exposto atingido pela intervenção são então comparados com os resultados normais ou médios em uma população maior, sem a intervenção, a qual o grupo exposto pertence ou é um caso de. “Um caso de...” é a expressão usada na ciência política. Os cientistas políticos que estudam um caso concreto em geral tentam interpretar isso como um exemplo de algo maior. Eles fazem isso para que seus casos possam elucidar algo mais geral, algo mais abrangente. Em controles genéricos, o objetivo é exatamente o contrário: o avaliador quer dizer algo mais particular sobre o mais específico.

A palavra *genérico* significa “relativo ou característico de todo um grupo ou classe”. Em inglês, *generic*, tem a mesma raiz de gênero (*gender*) e estilo (*genre*). Controles genéricos também são chamados de *critério de delineamento da população*². (MOHR, 1995, p.185).

Controles genéricos são preferencialmente aplicados quando a intervenção avaliada está situada em um nível inferior ao nível nacional, por exemplo, em um município. Se o programa é dirigido a crianças com menos de sete anos em um município específico, a mesma categoria de crianças da nação como um todo pode ser usada como a população de referência maior. Se as atividades são dirigidas a todos os cidadãos do município A vivendo em moradias individuais, a avaliação pode usar como um gênero maior, toda a classe, aqueles que vivem em moradias individuais em todos os municípios, exceto A, 1) no condado; 2) de alguns condados vizinhos; e 3) da nação.

Suponha que um município do país, como Tierp no condado de Uppsala, tenha vivenciado, há alguns anos, aumento de custos e receitas decrescentes em seu departamento de assistência para indivíduos e famílias. Portanto, o governo municipal decidiu lançar três programas: a proibição de aquisições suplementares, suspender novos projetos e um desmantelamento de projetos em curso. Depois de alguns anos um avaliador é contratado para realizar uma avaliação dos efeitos totais das três intervenções. Este avaliador decide aplicar controles genéricos, escolhendo comparar a evolução dos custos e dos rendimentos em Tierp por alguns anos, antes e após os três

² *Critério de delineamento da população* (comumente conhecidos como delineamento pareados ou combinados) consiste da comparação do grupo de controle com o grupo que recebe tratamento em variáveis que tenham algum efeito esperado sobre o *outcome*. Envolve a definição de *critérios de inclusão* de uma amostra da população nos grupos submetidos ao tratamento.

programas terem sido inaugurados com o mesmo desenvolvimento em cinco gêneros. Como “gêneros inteiros” ou casos de controle são usados? Algumas opções:

- 1) média consolidada para os municípios vizinhos Älvkarleby e Östhammar;
- 2) média consolidada para todos os municípios no condado de Uppsala, exceto Tierp;
- 3) média consolidada para todos os municípios em Uppsala (exceto Tierp), Västmanland, Örebro e Södermanland com estrutura semelhante a Tierp;
- 4) média consolidada para todos os municípios do país com estrutura semelhante a Tierp (excetuando-se Tierp) referente a estrutura de negócios, estrutura da população, etc; e
- 5) média consolidada para todos os municípios do país (excetuando-se Tierp), mas com o tamanho da população semelhante (20 000 - 29 999 habitantes).³

Controles genéricos decorrem da ideia de que o desenvolvimento no menor grupo abrangido pela intervenção deve ser comparado com os dados sobre o desenvolvimento médio ou normal para o mesmo período de tempo em uma ou várias populações maiores às quais o grupo menor pertence.

Um caso fornecido por Rossi e Freeman (1999, p. 331) são os programas de água fluoretada para reduzir as ocorrências de cáries dentárias. A distribuição aproximadamente normal de cáries dentárias na população dos Estados Unidos era bem conhecida. Quando experimentos com água fluoretada foram lançados em algumas comunidades foi possível, a partir disso, comparar a frequência de cáries nessas comunidades com a frequência normal na população como um todo. Graças a isso, uma medida aproximada dos efeitos do programa de água fluoretada foi alcançado, tendo todas as outras coisas permanecido iguais.

A partir deste caso podemos ver que controles genéricos implicam o uso de comparações entre casos. No entanto, os casos para comparação não são selecionados aleatoriamente nem combinados. Eles são determinados pelo *raciocínio argumentativo*.

³ Caso Tierp deva ser comparado aos municípios vizinhos Älvkarleby e Östhammar a comparação deve ser feita uma de cada vez para que fossem feitas correspondências no delineamento de controles.

Por isso, os avaliadores teimosos salientam que os controles genéricos devem ser usados com extrema cautela, examinando-se intensamente se os casos escolhidos para comparação são equivalentes aos participantes em todos os sentidos críticos.

Controles genéricos podem ser experimentais no sentido em que o avaliador pode distribuir a intervenção a um grupo menor. Normalmente, porém a intervenção é decidida e implementada por outra pessoa.

14.2 Controles Reflexivos

Similar aos controles genéricos, um pré-requisito para o delineamento de controles reflexivos é que uma série de medições seja realizada em intervalos consecutivos apropriados mostrando a pré-intervenção (linha de base), bem como as medidas pós-intervenção na variável resultado adequada. O delineamento de controles reflexivos difere dos controles genéricos levando-se em conta que nenhum caso de referência contemporâneo é usado. Apesar de ser um delineamento de um grupo só, controles reflexivos contêm um controle: o grupo de investigação é considerado como sendo o seu próprio grupo de controle. O termo *reflexivo* significa “referindo-se a si mesmo”. Marcação pré-intervenção servem como controles em relação às marcações pós-intervenção. Isto sugere que *deve* haver uma série temporal relativamente longa de marcações pré-intervenção ou dados de referência.

Controles reflexivos podem ser aplicados tanto em níveis agregados, bem como aos níveis individuais. Um grupo de investigação agregado é uma categoria em que os indivíduos pertinentes podem mudar ao longo do tempo. Tal unidade agregada pode ser de pessoas que vivem em habitações multifamiliares na cidade de Gotemburgo, moradores de rua na cidade de Gotemburgo, ou pessoas estrangeiras vivendo na cidade de Gotemburgo. Se observações repetidas em um delineamento de controles reflexivos dizem respeito ao nível do grupo agregado, a categoria como um todo é medida, embora os indivíduos que compõem a categoria entram e saem (ROSSI; FREEMAN; LIPSEY, 1999, p. 276)⁴.

⁴ Controles reflexivos aplicados aos indivíduos são muitas vezes chamado de *painel design* ou *painel longitudinal*. Os estudos de painel, em outras palavras, seguem sempre o mesmo grupo de pessoas, indivíduo a indivíduo ao longo do tempo, e não uma categoria agregada de pessoas, independentemente de quem são os indivíduos. (FITZ-GIBBON; MORRIS, 1978, p. 57; ROSSI; FREEMAN;

O delineamento de controles reflexivos é normalmente usado quando a intervenção é executada de forma abrangente. Desta forma, não há nenhuma maneira que os avaliadores possam usar algum tipo de grupo de controle simultâneo/contemporâneo separado, porque toda a população relevante é, em princípio, afetada pelo programa. Uma vez que esta situação é bastante comum no setor público, o delineamento de controles reflexivos vai ser, a partir deste ponto de vista, bastante útil. A lógica dos controles reflexivos foi sucintamente descrita por Rossi e Freeman (1985, p.299):

A justificativa essencial para a utilização de um delineamento de controle reflexivo é que, em algumas circunstâncias, é razoável acreditar que as metas permanecem idênticas de forma relevante, antes e depois da participação. Em outras palavras, em tais circunstâncias, pode-se supor que, sem a intervenção, os resultados das marcações pré-intervenção e pós-intervenção teriam sido o mesmo; portanto, se as alterações aparecessem após a intervenção, tais mudanças seriam diretamente atribuíveis à intervenção (ou seja, o impacto líquido seria igual ao impacto bruto).

Como todas as outras abordagens até aqui descritas, o delineamento de controles reflexivos é uma estratégia para indicar quais dados devem ser coletados, a fim de sustentar uma análise dos efeitos do programa, e não um método técnico para coleta de dados ou uma técnica de processamento de dados já recolhidos. No projeto de controles reflexivos, a atenção do avaliador é absolutamente focada em *um só grupo* - o grupo destinado a receber tratamento pelo programa - enquanto as variações na variável alvo são medidas (O = observação) em pelo menos uma ocasião antes ($O_{\text{pré}}$) e uma depois da adoção do programa X ($O_{\text{pós}}$).

LIPSEY, 1999, p.267, VEDUNG 1982a, p.112). Na continuação, o Delineamento de Série Temporal Interrompida e o Delineamento Antes-e-Depois para um Grupo concluirá a representação todos os delineamentos de controles reflexivos.

O delineamento de controles reflexivos assume diferentes formas. Dois deles serão elaborados aqui: Delineamento de Séries-Temporais Interrompidas e Delineamento de Antes-e-Depois para um Grupo. Ambos são amputações graduais do Delineamento Série Temporal de Controle, o mais forte, considerando uma perspectiva *ceteris paribus*, entre as estratégias de dois grupos quase-experimentais. A ideia de amputação deveria ser óbvia a partir da figura 14.1.

Figura 14.1
Dois Projetos de Controles Reflexivos:
Séries-Temporais Interrompidas e Antes e Depois para
um Grupo — como Quase Experimentos Amputados—

Delineamento	Atribuição	Observação pré-programa			Programa	Observação pós-programa		
<i>Delineamento Controles Reflexivos</i>								
Delin. Séries Temporais Interrompidas	Experim.							
Grupo de Investigação	Naturais	O ₁	O ₂	O ₃	X	O ₄	O ₅	O ₆
						Desenvolvimento contra-fatual estimado		
Delin. Antes-e Depois para um Grupo	Experim.-		O ₁		X		O ₂	
Grupo de Investigação	Naturais						Desenvolvimento contra-fatual estimado	
<i>Delineamento Quase experimental</i> (Delineamento Série de Controle)								
Grupo de Investigação	M	O ₁	O ₂	O ₃	X	O ₄	O ₅	O ₆
Grupo de Comparação	M	O ₁	O ₂	O ₃	---	O ₄	O ₅	O ₆

Experimentos Naturais = intervenções são examinadas à medida que vão se desenrolando na vida real, e não é feita nenhuma tentativa de adequá-las às exigências de experimentos aleatórios ou concordantes.

M = atribuição ou designação por concordância

O = observação (medição) do resultado (da variável resultado)

X = programa (intervenção, política, projeto, atividade, serviço, tratamento, input)

Fonte: O esquema é uma elaboração de uma imagem em Riecken e Boruch (1974, p. 98).

14.3 Delineamento de Séries-Temporais Interrompidas e Delineamento Antes-e-Depois para um Grupo

Em ambos os Delineamento de Séries-Temporais Interrompidas e Delineamento de Antes-e-Depois para um Grupo, nenhum grupo de comparação está incluído. Há apenas um grupo, o grupo de investigação. O Delineamento de Séries-Temporais Interrompidas incorpora medições periódicas repetidas (O_t) no mesmo grupo ao longo do tempo. Algumas medidas devem ser tomadas antes ($O_{1,2,3}$) e algumas após a intervenção ($O_{4,5,6}$). No Delineamento de Antes-e-Depois para um Grupo apenas uma pré-medição (O_1) e uma pós-medição (O_2) são tomadas. O objetivo da análise é, é claro, inferir se a intervenção (X) teve qualquer impacto. Se o fizesse, seria de esperar que as medidas pós-programa fossem diferentes a partir de observações pré-programa. As séries temporais devem mostrar sinais de uma “interrupção” depois que o programa é introduzido. Daí “interrompida” na designação⁵.

Por mais que não haja um caso de comparação, um desenvolvimento contrafactual é estimado em ambos os delineamentos. É calculado pelo *raciocínio argumentativo* a partir das marcações pré-intervenção do grupo de investigação. As marcações pré-intervenção de um único grupo de investigação são usados como ponto de partida para o raciocínio argumentativo cuidadoso por parte dos avaliadores de impacto para pensarem no que teria acontecido se a intervenção não tivesse ocorrido.

⁵ (MARK; HENRY; JULNES 2000, p. 265). Muitas vezes, em política pública, o grupo todo não pode ser estudado, nem mesmo com o Delineamento de Antes-e-Depois para um Grupo. Devido a restrições de tempo ou de dinheiro, por exemplo, pode ser impossível para coletar dados para toda a Suécia, a fim de esclarecer os efeitos de uma regulamentação nacional específica. Nestes casos, algumas amostras devem ser feitas, ou alguns casos devem ser selecionados. No entanto, isso não significa que o avaliador mudou para algum outro delineamento de pesquisa para esclarecer a questão da causalidade. O delineamento a ser utilizado é um delineamento de controles reflexivos. No entanto, além dos problemas de lidar com o delineamento de controles reflexivos adequado, o avaliador agora também terá alguns problemas de amostragem.

Quanto ao problema contrafactual em si, a ideia fundamental é que a linha de desenvolvimento observada nas marcações pré-programa deve continuar aproximadamente da mesma forma nas medidas pós-programa, a menos que o programa tenha sem permissão influenciado o padrão de desenvolvimento esperado. Com base em medições pré-programa (na figura 14.1 O_1 , O_2 , O_3 no Delineamento de Séries-Temporais Interrompidas, O_1 no Delineamento de Antes-e-Depois para um Grupo), os avaliadores de impacto formam uma expectativa em relação às prováveis marcações pós-programa ("desenvolvimento contrafactual estimado" na figura 14.1), se o programa não foi instalado. Esta linha contrafactual de desenvolvimento é comparada com o desenvolvimento real e a diferença pode ser atribuída ao programa. Quanto mais longo o período de tempo abrangido por medidas pré-programas, maior a capacidade dos avaliadores de prever a forma geral da trajetória contrafactual e, assim, a nossa capacidade de verificar o impacto do programa.

Ocasionalmente, os avaliadores de impacto podem prever com segurança o que teria acontecido se o governo não intervisse. Uma avaliação do programa de subvenções do governo sueco para pesquisa de energia solar sustentou que as células solares, os coletores solares, as instalações de armazenamento de calor e outras inovações técnicas que tinham sido desenvolvidas deviam ser resultado do programa. O argumento justificativo para isso foi que as inovações não eram competitivas no mercado em comparação com outros métodos de produção de energia; portanto, elas nunca teriam sido desenvolvidas, se o programa de subvenções não fosse adotado (EFN-rapport, 1985, p.13).

Naturalmente, delineamentos de controles reflexivos não podem ser baseados em dados estatísticos e documentos de texto apenas. Às vezes, os métodos de coleta de dados interrogativos - entrevistas e questionários - são adequados. Em alguns casos, a fotografia é um método muito bom para capturar qualidades e totalidades. Isso tem sido usado na política florestal sueca (ECKERBERG, 1987, p.35).

O que o Delineamento de Séries-Temporais Interrompidas tem a ver com o exemplo de execução do programa de Declaração de Combustível? Os suecos compraram carros mais eficientes em termos de combustível por causa do programa? Suponhamos que este tenha sido de fato o caso. Enquanto em 2010 o consumo de combustíveis totalizou uma média de 9,5 litros por 100 km e, em 2011, aumentou para 9,6 litros por 100 km, em 2012 caiu para 9,5 litros, em 2013,

para 9,4 litros, em 2014 para 9,2 litros, em 2015, para 9,1 litros, e em 2016 para 9,0 por 100 quilômetros. A declaração parece ter produzido o efeito pretendido.

Esta conclusão, no entanto, é prematura. A queda observada pode ser parte de uma tendência maior de veículos mais eficientes em termos de combustível. Esta tendência foi também, de fato, o caso. A propensão parece já ter começado no ano de 2012. O movimento em direção a carros menores era claramente perceptível bem antes das diretrizes entrarem na agenda política em 2014 e serem promulgadas a partir de 1º de Janeiro de 2015. Isso apoia a visão de que outros fatores além da Declaração de Combustível podem ter sido eficazes.

Isso ilustra a grande fraqueza do Delineamento de Séries-Temporais Interrompidas e de outros controles reflexivos. Estamos longe de ter confiança a respeito do efeito da intervenção pública. Ela pode ter tido um impacto, mas algumas outras mudanças - *fatores de externos de desordem* - podem ter produzido o desvio no resultado esperado nas medidas pós-programa. No caso da Declaração de Combustível, poderia ter sido um aumento de preços no petróleo bruto, ou as fortes advertências dos líderes políticos sobre a importância da economia de recursos energéticos escassos. As séries temporais no desenvolvimento de compras de carro não fornecem pistas confiáveis sobre os possíveis impactos desses *fatores de externos de desordem*.

No entanto, o Delineamento de Séries-Temporais Interrompidas oferece algumas possíveis explicações alternativas. Se uma pausa na tendência ocorrer muito tempo após a intervenção, podemos argumentar que o programa não teve qualquer impacto. Se as mudanças surgem bem antes de o programa ser instituído ou até mesmo antes de ter sido considerado, não podem ser a causa. Além disso, se as transformações ocorreram em conjunto com alterações numa outra contingência que sabemos que geralmente influencia as dimensões-alvo, isso pode fornecer uma base para a *discussão e argumentação* sobre o papel desse outro fator como causa provável.

O Delineamento de Séries-Temporais Interrompidas aponta, então, a importância de adquirir uma visão geral das forças de não intervenção que podem influenciar no resultado, em outras palavras, que trabalham com uma boa *teoria de situação* (*teoria do campo de intervenção*). Se a intervenção implica em persuadir os consumidores

através de informações a comprar carros mais eficientes em termos de combustível, então uma teoria de situação conteria as contingências mais importantes que influenciam a escolha dos novos carros pelas pessoas. Isso também indicaria como essa influência ocorre. Tal teoria relevante do campo da intervenção é necessária como uma ferramenta complementar ao delineamento de controles reflexivos. Estes extras fazem os projetos de controles reflexivos levar muito mais tempo e ser mais elaborados e trabalhosos. Com todas estas considerações, então realmente vale a pena utilizar controles reflexivos no setor público?

As compras de automóveis, provavelmente, são influenciadas pelo crescimento da renda disponível. Quanto maior for o aumento de renda disponível que as pessoas disfrutem, maiores serão os carros que provavelmente comprarão, até certo ponto. Outro fator influente é a evolução dos preços do combustível. O aumento dos custos de combustível pode levar os consumidores a comprar mais carros eficientes em termos energéticos. Além disso, a compra de carro, provavelmente, é influenciada por certas mudanças sociais bem gerais. Dois carros por família é um padrão cada vez mais comum. O segundo carro comprado é muitas vezes menor do que o primeiro, talvez devido ao fato de que as mulheres usarem este carro em particular mais do que os maridos e as mulheres tendem a preferir carros menores e mais versáteis. Ainda, outro fator é que as empresas cada vez mais compram carros e os concedem a seus executivos e gerentes. Isso pode levar a um aumento das compras de carros maiores. Estes são exemplos de contingências a serem incluídas em uma *teoria argumentativa* do campo de intervenção sobre compra de carros novos.

Nos *fatores de externos de desordem* não programados, deixamos o campo do delineamento para a coleta de dados e entramos na área de métodos de processamento de dados estatísticos.

Finalmente, uma advertência contra as interpretações exageradas das informações obtidas com o delineamento de controles reflexivo se justifica. Ele é suscetível a ameaças à validade interna. Em primeiro lugar, a causa da mudança pode ser o malévolos efeito Hawthorne. Se as marcações de medição são extraídas de fontes intrusivas como entrevistas, os entrevistados poderiam entender que os carros energeticamente eficientes são algo importante. Essa

percepção pode induzi-los a prestar mais atenção ao consumo de combustível do que de outra forma até a próxima situação de entrevista. Esta, por sua vez, pode induzi-los a mudar seu comportamento de um jeito que não poderia ser alcançado de outra forma.

Em segundo lugar, as variações podem ser devidas a modificações desafortunadas do grupo de investigação. A taxa de abandono dos indivíduos pode ter sido maior durante pós-medição que na pré-medição. E poderia ter sido tendencioso. Uma desistência grande ou enviesada entre a primeira e a última medição pode ter resultado na alteração registrada. E em terceiro lugar, o instrumento de medição pode ter sido ajustado durante a investigação, produzindo resultados imprecisos (RIECKEN; BORUCH, 1974, p.107).

O Delineamento de Séries-Temporais Interrompidas é altamente recomendado em situações em que variáveis-alvo são altamente voláteis e sujeitas a alterações. Quando as variáveis -alvo são estáveis ao longo do tempo, e as medidas reguladas pelos programas são relativamente fortes e abruptamente introduzidas, então o Delineamento de Antes-e-Depois para um Grupo mais simples, com uma medição pré-programa e uma medição pós-programa pode substituir o Delineamento de Séries-Temporais Interrompidas.

14.4 Controles-Sombra

No delineamento de controles-sombra, o efeito líquido de uma intervenção é estimado por pessoas com percepções especiais. *Sombra*, neste caso, significa “representação imperfeita e fraca” ou “imitação de alguma coisa”⁶.

Se os dados pré-intervenção e pós-intervenção das variáveis -alvo estão disponíveis, as pessoas com conhecimentos especiais são convidadas a contar o que teria acontecido se o programa não fosse instalado ou se os destinatários não fossem submetidos a atividades do programa. Em vez de realmente medir o caso contrafactual, o avaliador pede para outros que o estimem.

⁶ Controles-Sombra (*Shadow controls*) em Rossi, Freeman e Lipsey (1999, p. 356), mas também “avaliação crítica [do impacto]”, p. 268 et seq.

Figura 14.2
Controles-Sombra

Delineamento	Atribuição	Observação pré-intervenção			Intervenção	Observação pós-intervenção		
<i>Delineamento de Série Temporal Interrompido</i>								
Grupo de investigação	Experim. Naturais	O ₁	O ₂	O ₃	IU	O ₄	O ₅	O ₆
						desenvolvimento contra-fatual estimado		
<i>Delineamento Antes-e-Depois de um Grupo</i>								
Grupo de investigação	Experim. Naturais		O ₁		IU		O ₂	
						desenvolvimento contra-fatual estimado		

Experimentos Naturais = intervenções são examinadas que se vão desenrolando na vida real, sem ajustá-los às exigências de experimentos aleatórios ou combinados.

O = observação (medição) do resultado variável

I = intervenção (programa, projeto, política adotada, reforma, atividade, tratamento, prática etc.)

Controles-sombra podem ser divididos em avaliações de *experts*, julgamento do administrador das intervenções e julgamentos dos clientes⁷.

Aferição por peritos envolve o uso de autoridades externas para produzir o caso contrafactual. Aqueles que conhecem o campo muito bem demonstram o seu conhecimento em realizações reais e estão familiarizados com os resultados das intervenções na área são convidados a projetar com base neste histórico para estimar se um resultado real específico é maior ou menor do que o resultado normal de programas bem-sucedidos naquela área em particular.

⁷ É preferível a expressão "aferição por peritos" para a locução de Rossi-Freeman-Lipsey "avaliação pelos *connoisseurs* (conhecedores)", porque os *connoisseurs* - pessoas conhecedoras com gostos muito diferentes dos do consumidor médio - são usados principalmente para avaliar o valor de algum fenômeno indevidamente intrincado, e não relações causais entre contingências. Por conta disso, avaliações de especialistas também são diferentes das avaliações pelos pares apresentados acima como um modelo de avaliação, pois a contribuição mais importante deste último é a aplicação de critérios de mérito e padrões de desempenho, e não estimativas de relações causais. É também preferível "julgamentos de clientes" a "julgamento de participantes".

Rossi e Freeman (1999, p.359) listaram uma série de fontes de dados e questões a serem levantadas sempre que essas fontes de dados são consultadas na avaliação do programa. Com base em sua lista, foi construída uma visão geral com uma terminologia própria na figura de 14.4 em fontes de dados e perguntas que os especialistas podem considerar quando estiverem fornecendo estimativas do impacto da intervenção.

Figura 14.3

Controles-Sombra: Fontes de Dados e Questões Apropriadas

1. *MÉTODOS DOCUMENTAIS: arquivos administrativos.* Informação deve abranger:
 - a. tamanho da intervenção
 - b. tipo de intervenção pedida pelos clientes
 - c. experiências conflitantes com clientes
 - d. experiências pós-intervenção dos clientes
 - e. custos da intervenção por cliente que completa a intervenção
 - f. mudanças do cliente relevantes para os objetivos da intervenção, avaliados por medidas antes e depois
2. *MÉTODOS INTERROGATIVOS: Entrevistas com clientes (participantes).* Entrevistas com clientes atuais e antigos da intervenção podem levantar questões como:
 - a. recrutamento dos clientes
 - b. motivação dos clientes
 - c. satisfação do cliente com o processo de serviço e entrega do serviço
 - d. progresso dos clientes com respeito ao alcance dos objetivos da intervenção
 - e. o que teria acontecido se não houvesse intervenção
3. *MÉTODOS INTERROGATIVOS: Entrevistas com stakeholders no contexto de intervenção.* Entrevistas com autoridades locais, administradores das intervenções diferentes em disputa, e de importantes instituições locais, e indivíduos locais com poder ou representantes de poderosas instituições locais (ex.: autoridades políticas) deveriam abranger os seguintes temas:
 - a. o valor da intervenção
 - b. em que extensão a intervenção é vista como ajuda ou como ameaça à comunidade
 - c. interesse em continuar a intervenção quando o período de demonstração acabar
4. *MÉTODOS DE OBSERVAÇÃO: inspeção direta da operação de intervenção.* Intervenções que exigem trabalho ativo com os participantes (por exemplo, visitas domiciliares, sessões em sala de aula), o trabalho deve ser observado diretamente pelos peritos convidados

Nos *julgamentos dos administradores de intervenções*, se solicitam os pareceres dos burocratas *street-level* ou gerentes de intervenção sobre o que teria acontecido sem a intervenção. Nos *julgamentos dos clientes*, os destinatários (participantes) respondem à mesma pergunta (ROSSI; FREEMAN; LIPSEY, 1999, p. 361). Ambas as impressões são suscetíveis de serem fundamentadas na extensa experiência de campo. Por outro lado, ambos os grupos têm uma tendência a exagerar os impactos da intervenção, a fim de salvaguardar a existência continuada da intervenção.

Imagine uma situação em que uma empresa de pequeno porte no Norte da Suécia recebeu apoio do Fundo Estrutural da União Europeia, a fim de aumentar o emprego de 15 para 25 funcionários. Agora vamos supor que você recebe uma missão em que você é comissionado para avaliar o impacto deste auxílio. Suponha que você decide usar controles-sombra. Você marca um compromisso no local com o gerente de negócios e viaja para o lugar. Você já sabe que ele expandiu sua empresa de 10 a 25, mas agora você pode verificar isso pedindo e obtendo sua confirmação. Mas você também faz a pergunta contrafactual: o que teria acontecido no quesito “emprego na sua empresa”, se você não tivesse recebido auxílio da UE? O auxílio não mudou nada, responde o gerente. Teríamos expandido para 25 de qualquer maneira. Mesmo antes de sabermos sobre o Fundo Estrutural, planejamos expandir e tínhamos realmente tomado decisões para tal. Mas, então, ouvimos falar sobre os Fundos Estruturais da UE e decidimos nos candidatar. Mas teríamos expandido mesmo sem a subvenção da UE. Conclusão do avaliador: a subvenção da UE não tem efeito sobre o emprego neste caso. Lição aprendida: controles-sombra podem fornecer resultados satisfatórios com muito pouco trabalho de coleta de dados e em um curto espaço de tempo.

Os defensores de experimentação, controles genéricos e análise estatística têm sentimentos mistos sobre controles-sombra em avaliação de políticas públicas. Se não completamente descartados, eles são considerados como o último recurso, caso os controles experimentais, genéricos e reflexivos sejam inviáveis.

Isto é totalmente diferente das opiniões dos avaliadores naturalistas orientados qualitativamente, que recomendam fortemente aos avaliadores o diálogo e a comunicação com os clientes de intervenção e outras partes interessadas. Para eles, concepções,

crenças, motivos e julgamentos do cliente e do administrador são os tipos mais importantes de dados para avaliação de políticas públicas.

Em muitas ocasiões, os controles-sombra são a única alternativa prática disponível, e nem mesmo os economistas orientados quantitativamente (ou quantitativos) se abstêm de usá-los. Em um estudo sueco sobre o impacto da regulamentação de preços imposta pelo governo sobre as empresas individuais, Lars Jonung (1984, p.149), um notável economista e assessor do antigo governo liberal-conservador liderado por Carl Bildt (1991-1994), pediu às empresas que informassem qual seria seu nível de lucro, se não tivesse havido nenhuma regulamentação. A alternativa contrafactual foi verificada por meio do uso de controles-sombra, ou, mais especificamente, o julgamento dos clientes.

Os defensores da experimentação, controles genéricos e métodos estatísticos depositam pouca confiança no julgamento dos clientes. Aos clientes falta tanto a amplitude da experiência quanto a competência perceptiva necessária para fazer julgamentos razoáveis relativos ao caso contrafactual. Há também um risco óbvio de que eles serão parcialmente a favor ou contra o programa.

Os mesmos pontos de vista também dizem respeito aos julgamentos do administrador da intervenção. Reconhecidamente, os administradores têm o melhor conhecimento do objeto da avaliação (avaliando), sua execução e seu funcionamento no campo, uma vez que eles podem ter trabalhado com os problemas durante anos. No entanto, os administradores do programa não têm uma atitude apropriada de ceticismo em relação a seu próprio trabalho. Administradores preservam seus próprios interesses e querem passar a melhor imagem possível de seus projetos. Rossi e Freeman (1989, p.371) afirmam que:

a avaliação de impacto bem conduzida toma como hipótese orientadora que o projeto não tem nenhum efeito, uma postura que vai exatamente contra o princípio que deve guiar a administração de um projeto, ou seja, que a intervenção tem efeitos importantes sobre os participantes. Esperar que reles mortais mantenham ambas hipóteses, simultaneamente, não é realista.

Controles-sombra não podem garantir resultados estanques. No entanto, eles não devem ser absolutamente descartados. Muitas vezes, não há outras opções disponíveis. Através do diálogo e da deliberação do avaliador com os clientes e administradores, algum conhecimento contrafactual do caso pode ser obtido, mesmo que bruto e fraco. E isso ainda é melhor do que conhecimento nenhum. Controles-sombra podem fornecer, por meio do diálogo e da argumentação, um retrato do que uma intervenção atingiu. Outra vantagem importante é que eles podem dar alguma dica de que maneira mecanismos generativos - como intenções e motivos - conectam a intervenção aos seus efeitos dos resultados⁸.

⁸ O delineamento de caso de controle, também conhecido como estudo de caso referencial, é difícil de classificar. A partir de uma população, como habitantes de um município específico, um grupo (caso) é selecionado com base em dados sobre alguma relevante variável de resultado, por exemplo, "realizaram as medidas pertinentes à conservação de energia", tais como a instalação de janelas de três lâminas de vidro em suas casas individuais. Primeiro, uma população foi dividida com base nos dados sobre a variável pertinente de resultados, por exemplo, em um grupo que tem (o caso, a janela tripla instalada) e um grupo que não tem (controles, referentes) executado medidas pertinentes à conservação de energia, tais como a instalação de janelas de três lâminas em suas casas individuais. Em seguida, outro grupo (controles, referentes) é selecionado entre a população, que é semelhante ao grupo de casos em muitos aspectos relevantes, mas diferem dele na conta que seus indivíduos não instalaram janelas de três lâminas em suas casas isoladas. O avaliador, em seguida, examina os dois grupos de trás para a frente no tempo tentando rastrear possíveis fatores causais incluindo, o sistema de intervenção pública ou intervenção relevante. Será que é realmente o caso de aqueles que instalaram janelas de três lâminas também terem notado e tomado parte na intervenção municipal que foi adotada e implementada mais cedo enquanto que aqueles que não instalaram nem participaram nem tinham notado a intervenção? Ou é o caso da diferença de instalação ser atribuída a qualquer outra coisa que diferencia os dois grupos entre si? Dividindo a população em dois grupos, com base nas diferenças relativas a alguma variável de resultado (variável de efeitos em potencial) e não uma variável intervenção, faz o delineamento de controle um caso diferente de todos os outros na nossa lista. Então, a ideia é procurar as causas, incluindo a intervenção. Contrariamente a isto, todas as outras abordagens partem do que sabemos sobre a situação da intervenção; nós sabemos quem tem e quem não tem sido exposto a alguma intervenção. A principal tarefa aqui é descobrir se a intervenção produziu quaisquer efeitos nos resultados. No delineamento de caso de controle, busque suas causas; em todos os outros delineamentos, busque seus efeitos.

AVALIAÇÃO POR PROCESS TRACING E TEORIA DA IMPLEMENTAÇÃO

Uma única estratégia governamental pode envolver as atividades complexas e inter-relacionadas de vários níveis de agências governamentais e organizações, organizações privadas, associações profissionais, grupos de interesse e populações de clientes. Como essa profusão de atividades pode ser controlada e dirigida? Esta questão é o cerne do que veio a ser conhecido... como o “problema de implementação”.

Eugene Bardach

Quais *outcomes* são causados por uma intervenção? São causados direta ou indiretamente, parcial ou integralmente, a curto ou a longo prazo, dentro ou fora das áreas-alvo? Essa também é a questão básica em *process tracing*. Ainda assim, *process tracing* ataca o problema de efeitos de forma diferente dos delineamentos até agora tratados.

15.1 Avaliação por Process Tracing é Conduzida como Estudo de Caso

Em contraste com os RCTs (Randomized Controlled Trial=Experimentos Aleatórios Controlados), controles-sombra, controles genéricos e outras abordagens de causalidade, *process tracing* não emprega caso de comparação. Concentra-se em um único caso, com foco em todo o seu processo de implementação entre a intervenção e o *outcome*.

Nas décadas de 1960 e 1970, as definições de estudos de caso basearam-se em distinções entre o estudo de um e de dois ou mais casos de um fenômeno. Os estudos de caso foram caracterizados como estudos de um-*n*, em contraste com estudos comparativos de dois ou mais-*n*, ou mesmo de estudos estatísticos comparativos de *large-N*. A distinção sugere que a variação no número de casos

examinados é a disparidade mais relevante entre os delineamentos de estudo de caso e outros delineamentos para análise causal. Os métodos de *large-N* são sempre preferíveis quando dados suficientes estão disponíveis para investigação, como Arend Lijphart sugeriu em seu influente artigo há quase cinquenta anos (1971).

Process Tracing envolve uma perspectiva diferente. Um caso é definido como um exemplo de uma classe de eventos. O termo “classe de eventos” refere-se a um conjunto de fenômenos reais de interesse acadêmico, como programas governamentais de conservação de energia, subsídios governamentais como instrumento de políticas, subsídios governamentais à indústria como instrumento de política, subsídios governamentais para o setor de mineração como um instrumento de política, projetos de investigação do governo em oposição a políticas permanentes, cooperação entre agências governamentais sobre aconselhamento para indústrias inovadoras em seu início etc. Os casos são construídos a partir de vários pontos de partida devido ao interesse de aferição de comissários e de avaliadores (GEORGE; BENNETT, 2005, p. 21).

Em segundo lugar, *process tracing* adota uma abordagem holística e orientada ao contexto, observando casos (o que está sendo avaliado) em seus ambientes naturais, seja da perspectiva política, administrativa, social ou espacial. Isso significa que os contextos podem variar enormemente em *process tracing*, enquanto outros delineamentos tendem a mantê-los constantes.

15.2 Abrindo a Caixa-Preta

Process Tracing abre a caixa preta. Outras abordagens, como controles-sombra e controles genéricos, concentram-se em eventual covariação entre a intervenção (*input*) e os *outcomes* (ou *outputs*), mas deixam no escuro as ligações entre a intervenção e as atividades subseqüentes dos destinatários. Os acontecimentos entre a intervenção e os vários destinatários são tratados como um recipiente que não está aberto, portanto a analogia com uma caixa preta. *Process tracing*, por outro lado, abre a caixa preta acompanhando de perto os vínculos entre a intervenção e o *outcome* (*output*) em todos os seus tortuosos caminhos. Dessa forma, a metáfora da abertura caixa preta captura a terceira grande disparidade entre o *process tracing* e outras abordagens para a causalidade, seja experimental ou não experimental (LANGE, 2013, p. 48 et seq., p. 69, p. 118 et seq., p. 154, p. 177).

Pesquisadores estatísticos concordam que a correlação, por mais forte que seja, não implica causalidade (GEORGE; BENNETT, 2005, p. 21). Em contrapartida, a essência do *process tracing* é que os estudiosos querem ir além da simples identificação de correlações entre variáveis independentes (Xs) e *outcomes* (Ys). O estabelecimento de laços, vínculos e relações descobertos entre eles é uma importante tarefa do *process tracing*.

Uma forte correlação estatística foi encontrada entre democracia e paz, argumentam Beach e Pedersen no primeiro capítulo de seu livro sobre *process tracing* (2013, p. 1). No entanto, como saber se a democracia mútua foi a causa da paz entre duas nações? Como a democracia produz relações mais pacíficas? “Responder a essas questões requer que se desvende a relação causal entre democracia mútua e paz para estudar o *mecanismo causal* que interliga os dois [fenômenos no mundo real]”, concluem (BEACH; PEDERSEN, 2013, p. 1. Grifo nosso).

Aqui, o termo “mecanismo causal” entra no texto. *Process tracing* significa descrever os mecanismos causais. “*Descrição cuidadosa* é um dos fundamentos do *process tracing*”, observa Collier (2011, p. 823). No entanto, é uma descrição cuidadosa emoldurada em um ângulo particular, ele continua. *Process tracing* analisa as trajetórias de mudança por meio de conexões chamadas mecanismos. É documentado como histórias cronologicamente ordenadas ou narrativas estruturadas de uma perspectiva de mecanismos. (adaptado de COLLIER, 2011, p. 823; ver também WARWICK; NOLAN, 2014, p. 5, 16 et seq.).

Mas o que é um mecanismo? Definições abundam na literatura. Segue, em seção separada, ponderações sobre a questão.

15.3 Mecanismos

Segue a sugestão de algumas definições:

Stuart S Glennan: um mecanismo causal pode ser definido como “um sistema complexo que produz um *outcome* pela interação de várias partes” (GLENNAN, 1996, p. 52, também em BEACH; PEDERSEN, 2013, p. 1).

Jon Elster: as peças e as engrenagens¹ que levaram as relações [entre políticas públicas e *outcomes*] à existência (ELSTER, 2007, p. 36; também em SANDAHL; PETERSSON, 2016, p. 62).

Alexander L George e Andrew Bennett: *Process tracing* envolve tentativas de identificar o processo causal interveniente – a cadeia causal e o mecanismo causal entre uma variável independente (ou variáveis) e o *outcome* da variável dependente (GEORGE; BENNETT, 2005, p. 206; também em BEACH; PEDERSEN, 2013, p. 1)

Essas definições são abstratas e carecem excessivamente de conteúdo específico. Uma definição de Carol Weiss é mais objetiva. Ela argumenta que os mecanismos não estão incluídos na intervenção ou na política pública. Em vez disso, o termo refere-se às impressões das pessoas sobre a intervenção e à transformação dessas impressões em respostas. Afirmar Weiss:

O mecanismo de mudança não são as atividades do programa *per se*, mas a resposta que as atividades geram. Por exemplo, em um programa de aconselhamento contraceptivo, se o aconselhamento estiver associado à redução da gravidez, a causa da mudança pode ser o aconselhamento. Mas o mecanismo não é o aconselhamento, que é a atividade do programa, o processo do programa. O mecanismo pode ser o conhecimento que os participantes adquirem com o aconselhamento. Ou pode ser que a existência do programa de aconselhamento ajude a superar tabus culturais contra o planejamento familiar; isso pode desencadear uma mudança nas relações de poder entre homens e mulheres. Essas ou algumas das várias outras respostas cognitivas, afetivas e sociais podem ser os mecanismos que conduzem aos *outcomes* desejados. (WEISS, 1997, p. 46 apud SANDAHL; PETERSSON, 2016, p. 62).

¹ Ver livro publicado em português tratando do conceito de mecanismos. ELSTER, Jon. **Peças e Engrenagens das Ciências Sociais**. Rio de Janeiro: Editora Relume-Dumará. 1994.

Em outras palavras, os mecanismos não são componentes da intervenção. A intervenção dispara mecanismos. No entanto, raramente é a intervenção como um todo que os dispara; são alguns dos traços particulares da intervenção que funcionam dessa maneira, sob certos contextos específicos e circunstâncias.

Em *process tracing*, o avaliador mapeia cuidadosamente como algum determinante causal (para repetir: geralmente não a intervenção como um todo, mas algum ingrediente particular da intervenção) libera algumas interpretações, que, por sua vez iniciam algumas considerações, raciocínios e ações entre os receptores, que, por sua vez, trazem algum *outcome*. São essas interpretações, considerações, raciocínios e ações que constituem mecanismos.

A noção de mecanismos (ou *mecanismos generativos*) é recuperada do discurso sobre a *causalidade generativa*, como é formulada na chamada *avaliação realista*. A avaliação realista argumenta que as intervenções podem funcionar de forma diferente para pessoas diferentes e podem produzir *outcomes* variáveis, dependendo do contexto em que uma intervenção se desenrola. A mudança é produzida ou gerada por meio de tais mecanismos. As intervenções não funcionam simplesmente *per se*. Elas (ou suas partes) só funcionam se ativarem mecanismos atores em alguns contextos particulares. Essa compreensão generativa da causalidade é bastante captada pela seguinte citação: "...as intervenções não funcionam, são as interpretações dos envolvidos que produzem resultados" (PAWSON, 2013, p. 34).

As intervenções funcionam porque disponibilizam recursos para os participantes (receptores) que, por sua vez, alteram ou reforçam suas considerações, raciocínio e, posteriormente, suas ações (PAWSON, 2013, p. 33, DALKIN et al., 2015). Essas mudanças em raciocínios e em ações são o componente central do conceito de mecanismo na avaliação realista. A mudança não é, portanto, causada pela intervenção como tal, mas igualmente pelo participante. A mudança é causada pelo raciocínio alterado ou fortalecido e, posteriormente, pelo comportamento modificado do participante envolvido. Portanto, os mecanismos da avaliação realista concernem as percepções, as interpretações e as ações dos participantes (alvos, destinatários). Sem interpretações dos recursos fornecidos e subsequentes mudanças de comportamento ou fortalecimentos, não pode haver mecanismos.

Além disso, os contextos (circunstâncias circundantes) também estão envolvidos. Se um mecanismo será ou não ativado e em que grau dependerá do contexto específico em que a intervenção se desenrola. Se os mecanismos não estiverem ativados, os *outcomes* desejados não serão alcançados. Por esse motivo, os contextos são importantes.

15.4 Process Tracing como Teoria da Intervenção

Na avaliação do setor público, um ponto importante do *process tracing* é testar empiricamente a razoabilidade da teoria da mudança inerente de uma intervenção. O primeiro passo é reconstruir, ordenar e esclarecer a crua e bruta teoria da mudança já inserida na intervenção. Para esse fim, o modelo de sistema estendido é usado como ferramenta. O modelo de sistema estendido é uma versão ampliada e alongada do simples **input** → **conversão** → **output** de modelo de sistema, modificado e magnificado para **input** **conversão** **output** → **alvos** → **outcome 1** → **outcome 2** → **outcome 3** etc, a fim de melhor atender ao estudo de processos do setor público. Interpretada e organizada nas categorias do modelo de sistema estendido, a teoria da intervenção refinada é então usada como uma ferramenta a seguir o caminho das ações originalmente estimuladas pela intervenção. *Process tracing* significa que o avaliador identifica cuidadosamente as marcas-chave, as pegadas e os eventos na cadeia de implementação (Figura 15.1), a fim de, gradualmente, ao longo do tempo, encontrar evidências dos efeitos de diferentes tipos aos quais contribuíram.

Com mais detalhes, o avaliador tenta ver quais ingredientes da intervenção costumam desencadear os raciocínios e ações dos destinatários nos vários níveis de implementação em seus contextos reais e naturais e como isso é interpretado, fundamentado e atuado por outros atores em outros níveis. Em *process tracing*, tenta-se adotar, reivindicar e representar ligações causais entre traços de intervenção, por um lado; e, por outro, as interpretações, considerações, raciocínios e ações subsequentes de vários destinatários, como na Figura 15.2.

A fim de fundamentar ou falsear a direção e a extensão em que uma intervenção realmente influencia os *outcomes*, o desdobramento temporal dos processos desencadeados pela intervenção – incluindo atividades repetidas de implementação, respostas imediatas, intermediárias e de grupos-alvo a essas atividades repetidas,

bem como fatores no contexto mais amplo – é cuidadosamente estudado, descobrindo e estabelecendo fatores e mecanismos generativos que, ao menos até certo ponto, foram agentes causais (LANGE, 2013, p. 48 et seq.).

O *process tracing* é usado para esclarecer a direção e a magnitude das sequências causais ao longo do tempo entre a intervenção e alguns *outcomes*. Ele pode identificar e representar janelas de oportunidade, momentos de formação (*formative moments*), pontos de inflexão e avanços no processo e isolar itens específicos que funcionam e não funcionam. Além disso, pode indicar não linearidades “nas quais pequenas ações iniciais ... reverberam de forma inesperada e imprevisível para ter impactos enormes” (PATTON, 2011, p. 193).

O uso de *process tracing* complementa outros delineamentos não processados, ao oferecer uma maneira de descrever em detalhes cronológicos os traços de intervenção que influenciam os mecanismos dos destinatários, ou seja, o raciocínio e as ações concomitantes induzidas por ele, fazendo com que outros raciocinem e ajam (BENNETT, 2010; COLLIER, 2011; PAWSON, 2013, p. 115).

Nesse caso, o avaliador inicia sua análise com a intervenção e procura por seus efeitos no futuro (*downstream process tracing*). Contudo, o avaliador também pode começar com os *outcomes* (efeitos brutos, efeitos potenciais) e, em seguida, procurar retroativamente por causas de bloqueio, a fim de determinar se eles podem ser seguidos até a intervenção (*upstream process tracing*).

15.5 Sete Fatores Amplos que Podem Influenciar os Outcomes da Intervenção

Sete fatores amplos, que podem afetar a implementação da intervenção pública, dados os contextos apropriados, estão dispostos nas figuras 15.1 e 15.2, graficamente. São enumeradas possíveis contingências – fatores de desencadeamento de mecanismo, determinantes de disparadores de mecanismos (*mechanism-triggering factors, mechanism-firing determinants*) – que podem contribuir para a consecução dos objetivos de *outcomes* e de *outputs* de intervenção e efeitos colaterais positivos, mas também dificultar o atingimento de metas ou desviar o processo para efeitos colaterais negativos. A ideia é que esses sete grandes fatores são de particular interesse no *process tracing*.

O esquema de sete fatores é o resultado de simplificações importantes. Presume-se que, em primeiro lugar, o quê está sendo avaliado – “avaliando” (*evaluand* em inglês) – – é uma intervenção pública adotada, o que é razoável, uma vez que este é um livro sobre avaliação do setor público. Uma intervenção pública adotada é o assunto a ser implementado, o quê será executado, a coisa a ser conduzida. Em segundo lugar, é uma intervenção no nível mais alto, central, de um Estado nacional, não uma intervenção global das Nações Unidas ou de alguma agência das Nações Unidas como a UNESCO, nem uma intervenção da União Europeia, nem um programa de nível municipal dentro de um Estado nacional.

Sobre os quatro principais fatores de implementação temos:

- 1) Implementação 1: a intervenção propriamente dita, que deve ser implementada (o “implementando” (*implementand* em sueco));
- 2) Implementação 2: implementadores superiores e intermediários;
- 3) Implementação 3: operadores de nível de rua; e
- 4) Implementação 4: membros do grupo-alvo.

Na realidade, os níveis de implementação proliferam no mundo moderno. No entanto, por razões pedagógicas e para não fazer o capítulo tão complicado e incômodo, condensou-se seus números centrais a quatro.

Alguns escritores incluem implementação na intervenção. Esse não é o caso aqui. As implementações 2, 3 e 4 são vistas como separadas da intervenção a serem implementadas e consideradas como parte da paisagem onde a intervenção aterriza. A implementação é de interesse óbvio para *process tracers*, uma vez que pode conter *drivers* (agentes causais) cruciais que afetam *outputs* e *outcomes*. Para destacar isso, serão tratados aqui como fatores por direito próprio, separados da intervenção em foco.

Os quatro principais fatores de implementação estão envolvidos pelos seguintes três tipos de fatores de contexto:

- 1) Contexto 1: formação da intervenção;
- 2) Contexto 2. intervenções paralelas, outras partes interessadas, cidadãos, mercados;
- 3) Contexto 3: monitoramento e avaliação *ex post*.

Alguns escritores incluem o processo de formação da intervenção, assim como seus atores e atividades envolvidos na intervenção a serem avaliados. Esse não é o caso aqui. O processo de formação e os atores que promovem o desenvolvimento e a adoção da intervenção não são considerados como se fizessem parte da intervenção propriamente dita, mas pertencem à paisagem a partir da qual a intervenção se origina (*landscape from which the intervention takes off*). Em outras palavras, a formação de intervenção é tratada como um ambiente (*environment*): Contexto 1: formação da intervenção.

Novamente, isso se deve ao fato de que a fase de formação pode conter agentes causais cruciais que afetam os *outputs* e os *outcomes* da intervenção. Para destacar isso, serão tratados aqui como determinantes por direito próprio, separados da intervenção em foco.

A Figura 15.1 também lista uma série de subfactores sob os principais sete fatores amplos. Para evitar fazer o capítulo muito longo e pesado, foram listados apenas alguns dos candidatos possíveis e relevantes. Os subfactores escolhidos devem ser tomados apenas como ilustrações limitadas. Para listas extensas de subfactores, o leitor deve consultar a grande literatura de implementação (leitores de inglês e de línguas nórdicas podem estudar, respectivamente, por exemplo, Vedung (1997) e Vedung (2016) para listas e referências pertinentes).

No melhor dos mundos das políticas públicas, presume-se que as atividades nos níveis 2, 3 e 4 sejam coordenadas entre si e com as inclinações de governança originais no nível 1 e, alinhadas com os indicadores de metas e de cronogramas no nível 1 também. Todo o sistema é simplificado para que tudo do sistema mais abaixo seja direcionado para a implementação perfeita dos objetivos e dos desejos mais substantivos e orientados para o processo.

Uma vez que a Figura 15.1 contém apenas uma lista, pode parecer absurdo chamá-la de uma teoria da implementação como é feita no cabeçalho deste capítulo; as expressões mais modestas “quadro teórico” ou “estrutura” seriam mais apropriadas. No entanto, é algo mais do que uma lista detalhada pela qual os *process tracers* devem procurar relacionamentos de conexão entre os fenômenos da lista. Na verdade, uma contingência pode tornar-se operacional somente em combinação com outras contingências facilitadoras. Por exemplo, o apoio político sob a forma de uma ampla frente partidária (5.1) pode produzir uma reforma difusa (1.3) que, por sua vez, pode

afetar a compreensão das agências nacionais ou a sua vontade de implementar (2.1), que no final irá afetar os resultados.

Em segundo lugar, a lista também contém referências a várias teorias sociais e administrativas: teorias de *path dependency* (1.1), empresários e campeões de produtos (2.2), a tríade compreensão-disposição-capacidade (3.1, 4.1-4), as estratégias de enfrentamento da teoria de burocratas de nível de rua (3.2), momentos formativos (janelas de oportunidade, conjunturas críticas 4.1.), teoria de captura regulatória (4.2), antecipação ("à sombra do futuro", 5.2), a teoria conglomerada da administração pública, reivindicação de crédito (7.1) e evitar a culpa (7.2).

Deve ser óbvio que os fatores listados na Figura 15.1 não devem ser lidos de modo preditivo. Eles não devem ser redigidos como, por exemplo, "se x faz k , então y irá ocorrer". A ambição é mais despretensiosa: apresentar uma lista de fatores (condições), que podem ser usados por avaliadores que usam *process tracing* para descobrir, após o fato, quais contingências deram origem a mecanismos que realmente contribuíram para os efeitos registrados do *output* e do *outcome* de alguma intervenção pública avaliada².

Figura 15.1 - Process Tracing: Fatores de Consolidação

1. IMPLEMENTAÇÃO 1: EFETUANDO FATORES NA INTERVENÇÃO
 - 1 Direção da mudança pretendida (*path dependency*/inércia)
 - 2 Agendas ocultas
 - 3 Clareza (obscuridade linguística, diversas opções de ação)
2. IMPLEMENTAÇÃO 2: RESPOSTAS DE NÍVEL ELEVADO E INTERMEDIÁRIO
 - 1 Agências nacionais: compreensão, capacidade, disposição
 - 2 Empreendedores, zelotes, campeões de intervenção (ou produto)

² Compare com Hargrove, (1975) e (1983); Pressman; Wildavsky (1984); Van Meter e Van Horn (1975); Montjoy e O'Toole (1979); Nakamura e Smallwood (1980); Allison, 1980; Berman (1978) e (1980); Winter (1990); Sabatier; Mazmanian (1980) e (1981); Elmore (1980); Mazmanian; Sabatier (1981); Municio (1982); Mayntz (1980) e (1982); Stone (1980) e (1985).

3. IMPLEMENTAÇÃO 3: RESPOSTA DO OPERADOR DE NÍVEL DE RUA

- 1 Compreensão, capacidade e disposição
- 2 Estratégias de enfrentamento

4. IMPLEMENTAÇÃO 4: RESPOSTA DO GRUPO-ALVO

- 1 Disposição/Capacidade: momento formativo (janela de oportunidade, conjuntura crítica)
- 2 Disposição: captura
- 3 Disposição: camuflagem
- 4 Disposição: free riders (efeitos do peso morto, *zero resource additionality*)

5. CONTEXTO 1: FORMAÇÃO DA INTERVENÇÃO E A SOMBRA DO FUTURO

- 1 Participação de interesses afetados na formação da intervenção
- 2 Antecipação (na sombra do futuro, na sombra da intervenção)

6. CONTEXTO 2: INTERVENÇÕES PARALELAS, OUTROS STAKEHOLDERS, CIDADÃOS, MERCADOS

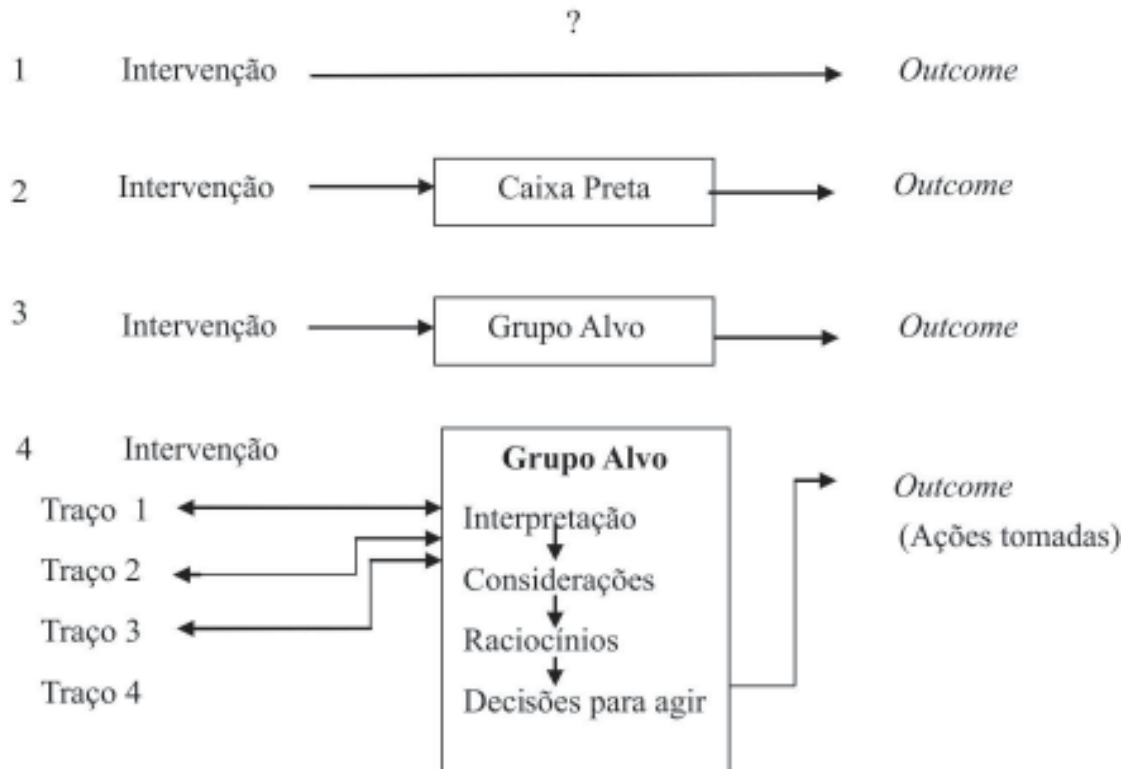
- 1 Investigações paralelas – individualmente ou em conglomerados
- 2 Desobediência civil
- 3 Mercados por meio de demanda de consumidores e mudanças de preços

7. CONTEXTO 3: MONITORAMENTO E AVALIAÇÃO EX POST

- 1 Monitoramento informal: demanda de crédito
- 2 Monitoramento informal: evasão da culpa
- 3 Monitoramento/avaliação formal

A representação das ligações causais entre traços de intervenção, por um lado; e, por outro, as interpretações, considerações, raciocínios e ações subsequentes de vários destinatários foi esquematizada graficamente na Figura 15.2.

Figura 15.2 – Process Tracing – Fatores de Consolidação Graficamente



Agora, serão apresentados os vários fatores de intervenção e contexto, um a um.

15.6 Implementação 1: Fatores Efetivos na Intervenção

Alguns traços ou características da intervenção podem ter um peso significativo para sua implementação. Pode-se exemplificar isso por meio destes três subfatores: *path dependency*/inércia, agendas ocultas e clareza. Na verdade, existem vários outros possíveis determinantes; mas, neste capítulo, serão abrangidos três apenas, para evitar o alongamento excessivo.

Primeiramente, **dependência da trajetória / inércia**. O grau de alinhamento ou desvio do conteúdo da intervenção corrente em relação a políticas anteriormente adotadas pode afetar a sua implementação e, conseqüentemente, também os resultados (*outcomes*). Se a mudança pretendida é considerada um passo natural na direção das políticas tradicionais os beneficiários e os agentes públicos vão realizar a intervenção mais rigorosamente do que se ela se constituir em um desvio ou em uma ruptura, as demais coisas permanecendo iguais.

Inércia refere-se à tendência de um corpo em repouso permanecer assim; ou se um corpo estiver em movimento continuar a mover-se em linha reta em velocidade constante salvo se uma força lhe for aplicada. De modo geral, inércia significa a aversão ao movimento, à ação, à mudança.

Se um movimento político perseverante e com grande força de vontade dominar o governo por um período de tempo longo, os servidores civis se acostumarão a trabalhar com certos tipos de problemas, instrumentos corretivos e arranjos organizacionais. Padrões estabelecidos de pensamento e hábitos enraizados irrefletidamente se instalarão. Se então uma mudança em política pública é lançada, por exemplo, por um governo com uma nova maioria obtida nas urnas, os novos detentores do poder vão descobrir que os burocratas não são facilmente manobráveis na nova direção. Implementadores serão lentos ou mostrarão sua preguiça nas ações em respostas às mudanças pretendidas pelo novo governo. Podem surgir dificuldades de ajustamento que poderão transformar-se em obstáculos à implementação.

Constatou-se que o governo radical conservador de Thatcher, que chegou ao poder, no Reino Unido, em 1979, teve problemas para reorientar os servidores civis³. Um caso semelhante ocorreu na Suécia em 1976, quando um governo não socialista chegou ao poder após 44 anos de quase ininterrupto domínio do Partido Operário Social-Democrata da Suécia. Suas dificuldades em reorientar a burocracia são um tema importante nas análises desse período entre 1976 e 1982 (RYDÉN, 1983)⁴.

³ Um olhar atento em um exemplo de dependência da trajetória relativamente à tecnologia é o teclado QWERTY, que não estaria em uso hoje se não tivesse sido escolhido há cem anos. Um exemplo singelo sueco é a instituição do tráfego rodoviário pela mão esquerda durante a Grande Guerra do Norte pelo Rei Charles XII em Bender, no então Império Otomano, a qual, com o tempo, drenou imensos recursos e custos para mudar para o tráfego rodoviário de mão direita em 1957.

⁴ A noção de inércia burocrática tem tido um papel significativo no pensamento socialista mais antigo no assunto relativo assunção de poder no estado burguês. Se os socialistas chegaram ao poder através de eleições ordinárias, servidores civis, acostumados a políticas liberais, orientadas ao mercado, obstruíam passivamente ou ativamente os novos governantes. Socialistas encontrariam grandes dificuldades na implementação das suas reformas e intenções. No longo prazo, eles poderiam render-se e tornar-se prisioneiros do estado burguês, se argumentava.

Inércia está relacionada com *dependência da trajetória*, que é a tendência de que os atores subordinados deixem as práticas do passado prosseguir mesmo que alternativas mais razoáveis estejam disponíveis. Imaginem uma decisão de política pública em um ponto A que requer uma nova trajetória. Para a administração isso significaria uma reorientação. E tal mudança de trajetória pode ocorrer, mas tão somente com indolência e lentidão. Por quê? Os pesquisadores mostram três possibilidades admitidas ou mecanismos de ação da dependência da trajetória (BENGTSSON et al., 2006, p. 25 et seq.). A Administração continua a perambular pela velha trajetória porque: 1) resulta em benefícios maiores e custos menores do que a nova trajetória, uma vez que uma mudança de sistema incorreria em tremendos custos de transação (relacionados com eficiência); 2) tem uma maior aceitação entre os próprios atores e o seu ambiente (relacionado com legitimidade); e, finalmente, 3) angaria maior apoio entre os atores dominantes (relacionado com poder).

As **agendas ocultas (política simbólica)** é uma segunda componente da intervenção a ser observada no *process tracing* avaliativo. Agendas ocultas envolvem o fato de que a intervenção é inaugurada para alguns outros objetivos acobertados ao invés de objetivos substantivos abertamente declarados. O propósito de objetivos divulgados e declarados publicamente pode ser para cobrir os objetivos reais que, por alguma razão, são julgados danosos ou embaraçosos se revelados. Ao revelar objetivos, intenta-se realizá-los; mas os objetivos ocultos podem ser os centrais.

Agendas ocultas são, muitas vezes, caracterizadas estrategicamente. **Considerações estratégicas** referem-se a crenças mantidas por atores pertinentes sobre como as possíveis escolhas de alternativas para ação vão impactar sobre escolhas de outros atores, e como isso, por sua vez, vai impactar sobre as suas posições atuais e futuras. Escolhas importantes feitas por um ator são precedidas por reflexões concernentes ao impacto dessas escolhas sobre escolhas de outros atores, o que, por sua vez, vai impactar sobre o posicionamento do mesmo ator. É a existência dessas considerações estratégicas por detrás das intervenções públicas que podem influenciar a sua implementação.

A política pública pode ser promulgada para dar a impressão de estar-se preocupado substantivamente com algo, sem necessariamente estar (LUNDQVIST, 1987, p. 127 et seq.). Vejamos o que disseram dois autores sobre o tema:

Nesse sentido a política pública pode tornar-se uma substituta para a ação, a fim de demonstrar que alguma coisa está sendo feita, sem atacar o problema real [...] governos ou formuladores de políticas querem responder, sem necessariamente querer tomar responsabilidade pela intervenção” (BARRETT; FUDGE, 1981, p. 276).

Formuladores de políticas querem ser lembrados por estarem firmemente no comando de uma situação, de modo a satisfazer opiniões partidárias ou reforçar a sua própria liderança partidária, manter a associação de filiados alinhada, conseguir votos em eleições, facilitar as coalizões governamentais e conseguir publicidade favorável na mídia. O propósito consciente pode também frustrar o moral dos membros de partidos competidores, para erodir sua força em eleições gerais, ou para enfraquecer a propensão de formar coalizões governamentais. Por essas razões, assim como em um jogo, os executores das intervenções podem achar vantajoso fingir ter um interesse no conteúdo substantivo da intervenção enquanto dissimula os propósitos reais, estratégicos.

Naturalmente, os servidores públicos das agências de níveis superiores e intermediários, assim como outros implementadores, podem perceber que as considerações estratégicas são as mais importantes e, por essa razão, devotam menos energia para atingir os objetivos substantivos do que teriam, de outra forma, feito (SAETREN, 1983; MARCH; OLSEN, 1976).

Um terceiro subfator de intervenção a observar no *process tracing* é **clareza**. Se, por um lado, a clareza é a vida e o sangue da ciência e da pesquisa; por outro a obscuridade está no centro da política. A obscuridade na intervenção **pode obstruir**. Objetivos opacos e turvos (EDWARDS III; SHARKANSKY, 1978, p. 297 et seq.; NAKAMURA; SMALLWOOD, 1980, p. 32 et seq.; VAN HORN; VAN METER, 1977, p. 108 et seq.; WILSON, 1989, p. 34-36) tornam difícil para os implementadores formar um quadro razoável de o quê as políticas públicas são delineadas para atingir; conseqüentemente, eles não podem apontar exatamente as discrepâncias entre objetivos declarados e *outputs* e *outcomes* reais, de modo a fortalecer a conformidade. Obscuridade permite aos implementadores o poder discricionário de acrescentar ou de subtrair ações infelizes à intervenção como concebida originalmente.

Catálogos de objetivos podem ilustrar essa infelicidade da obscuridade. Introduzida em 1975, a Lei Sueca da Construção estabelece em seu artigo 136:

O surgimento da importância de economizar energia nas atividades industriais e similares, nos recursos de fibras de madeira, e nos recursos de água e terra colecionados pelo país devem ser aprovados pelo governo” (*Government Bill 1 e 2*).

De acordo com os documentos da história legislativa da cláusula, a autoridade que concede as permissões precisa equilibrar a economia de energia com a economia em recursos de fibras de madeiras e os suprimentos de água e de terra da Suécia. Além desses quatro objetivos econômicos, outros objetivos deveriam ser considerados, como “objetivos sociais e [ambientais] que são importantes no planejamento de uso do solo, incluindo a conservação de recursos naturais escassos, o emprego estável e de alto nível, o equilíbrio regional, e o crescimento econômico”. (*Government Bill 1 e 2*). Em contrapartida, nunca foi clarificado como criar esse contrapeso entre oito objetivos diferentes. Resultado: é impossível conhecer exatamente qual *outcome* a coalizão governante queria atingir.

Ainda assim, a obscuridade na intervenção **pode** também **promover, estimular ou fomentar**. A escuridão verbal pode pavimentar a via para a criatividade e para a inovação, além de assegurar apoio de um grupo importante na implementação. A escuridão pode facilitar a coordenação de níveis mais baixos, a criação de redes e de novas atividades, que podem, por sua vez, criar *outputs* e *outcomes* que nenhum tomador de decisões de alto nível pode realmente antecipar (STONE, 2013, p. 178 et seq.).

Ocasionalmente, políticos não conseguem reunir tempo suficiente, intuições e habilidades para penetrar cuidadosamente em todos os assuntos que competem simultaneamente por sua atenção. Percebendo isso, os políticos expedem mandatos ou políticas públicas amplas, que, em termos gerais, declaram alguma direção, a fim de prover às agências regulatórias, aos intermediários, e às redes complexas de atores alguma margem de manobra para desenvolver interpretações mais específicas. Isso é **obscuridade como uma**

estratégia razoável da intervenção. Paradoxalmente, a linguagem geral e indeterminada da intervenção pode criar precondições para a acurácia e a precisão adaptada para cada situação avançada na implementação. Indeterminação (“entre feno e grama”, “entre falcão e urubu”) pode resultar na realização mais fidedigna dos objetivos da intervenção do que seria o caso se tivessem sido muito claros e específicos, necessitando uma enorme coleção de dados pertinentes para que pudessem ser promulgados. Nesses casos, a indeterminação da política pública muda a responsabilidade para emitir diretivas específicas dos tomadores de decisão seniores aos implementadores formais, às organizações de interesses afetadas e às redes de atores linha abaixo.

Um caso especial é a governança por regulação estatutária e regulamentação administrativa. Nesse caso, a intervenção primária é formulada pelos parlamentos como uma ampla lei estatutária que indica uma direção geral. Além disso, os parlamentos elaboram procedimentos legislativos para o contínuo aperfeiçoamento dos conteúdos estatutários. Ele habilita uma agência reguladora do governo (ou um governo) para cumprir essa tarefa. (WEST, 1985; ou BREYER, 2009; ESPING, 1994; FUNK; SEAMON, 2006).

Nesse momento, algo diferente deve acontecer. A agência reguladora continuará o esforço da organização, estabelecendo grupos mistos de burocratas / stakeholders para o desenvolvimento da regra (norma). Os resultados disso são, então, desenvolvidos por audiências ou por procedimentos escritos remissos. O resultado é uma série de regras ou normas específicas.

Esse é o procedimento regulamentar prevalente nas democracias ocidentais em matéria de saúde pública, ambiente de trabalho, uso do solo, construção de moradias e proteção ao consumidor.

Mais uma vez, isso é indeterminação como uma estratégia política razoável ou racional. A obscuridade promove, estimula, fomenta. Por meio da obscuridade no nível superior, a governança de nível inferior será mais precisa e clara. Além disso, a criação de regras administrativas pode aumentar a legitimidade, ao permitir que as partes interessadas participem da criação de normas, o que, por sua vez, pode aumentar a eficácia final.

Aparentemente, a implementação, nesses casos, é uma atividade criativa, um empreendimento construtivo e inovador, um

exercício de especificação de regras contínuas, embora sob os amplos limites (um guarda-chuva) de algum estatuto fundamental. Pode ser considerado como uma formação alargada de políticas, e não uma execução automática de metas e decisões pré-estabelecidas⁵.

A indeterminação também pode promover a governança colaborativa (cocriação)⁶. Se a intenção é promover a coprodução entre várias agências públicas, clientes / alvos e partes interessadas da sociedade civil, os partidos governantes podem tornar a intervenção intencionalmente ampla e flexível, a fim de fornecer espaço aos agentes interessados mais abaixo nas hierarquias, dando asas à sua criatividade e garantindo importante suporte de um grupo na implementação. Os objetivos parecidos com um guarda-chuva emitidos no nível superior podem encorajar a criação de metas, instrumentos de política e arranjos organizacionais mais específicos no nível inferior imediato, o que, ao ser implementado nos níveis mais baixos, pode aumentar a acuidade e a precisão para a população-alvo, de uma forma que o nível superior nunca teria sido capaz de prever e de criar (VEDUNG; BJURULF, 2015; HERTTING; VEDUNG, 2009, p. 48 e 49).

15.7 Implementação 2: Respostas dos Níveis Superior e Intermediário

O segundo fator de implementação principal a considerar em *process tracing* são as atividades executadas pelos atores nos níveis superior e intermediário. Os implementadores situados no nível alto tomam consciência dos recursos da intervenção e dos sinais de governança, interpretam-nos e determinam uma linha de ação dirigida aos implementadores de nível intermediário imediatamente logo abaixo de si. Estes, por sua vez, atuam da mesma forma com os outros atores intermediários abaixo na cadeia de comando, ou seja, os operadores de nível de rua.

⁵ Sobre objetivos indeterminados e obscuros, mas orientadores de uma direção, mobilizadores e amplos como um guarda-chuva, ver Hertting (2015, p. 192 et seq.); Hertting; Vedung (2009, p. 49) e Sahlin-Andersson (1989) (estratégia da obscuridade).

⁶ Também Nova Governança Pública ou New Public Governance, no original, (não NEW Public Government NPM), Krogstrup (2017, p. 33 et seq.); Osborne (2006); Xu; Sun; Si (2015); Pestoff et al. (2012).

Recomenda-se o uso da tríade *compreensão-vontade-capacidade*, recuperada de Lennart Lundqvist, em *Implementation Steering* (1987, p. 76 et seq.), para *process tracers* na busca de fatores impactantes. Idealmente, os *process tracers* devem fazer três perguntas sobre implementadores de nível superior e intermediário (e, claro, outros implementadores de nível inferior também – mais sobre isso mais tarde).

1) *Compreensão*; em que medida os implementadores *tem domínio* da aceitação pretendida da intervenção?

2) *Vontade*; em que medida os implementadores *querem* realizar o curso de ação, as atividades, contidas ou implícitas na intervenção?

3) *Capacidade*; em que medida os implementadores possuem habilidade ou capacidade (recursos) para realizar o curso de ação, as atividades, contidas ou implícitas, na intervenção? (GJELSTRUP, 1991, 19 et seq. ALEXANDERSON, 2006, 143 et seq., 237 et seq).

A comunicação de palavras e significados tem um papel tremendo na implementação de políticas públicas. Diferentes termos e as frases destacam-se em contextos que mudam constantemente, o que pode criar problemas de interpretação para implementadores de alto nível, bem como executores mais adiante nas cadeias de implementação.

Um exemplo pode ajudar a entender isso melhor. Às vezes, o Estado descarta uma doutrina dominante de administração e adota outra. No setor escolar, no início da década de 1990, abandonando o gerenciamento orientado a processos, o governo sueco introduziu o gerenciamento por resultados. Ao mesmo tempo, o sistema escolar público foi municipalizado. Vários estudos mostram que os implementadores intermediários, como os membros eleitos nos comitês executivos municipais e os comitês escolares, não usaram o amplo espaço de manobras fornecido pelo novo sistema de gestão. Embora sabendo que um novo sistema de gestão havia sido introduzido, os funcionários pertinentes não perceberam detalhadamente até que ponto eles agora podiam decidir por conta própria sobre os objetivos intermediários e as tecnologias, os métodos e os meios de intervenção para atingir esses objetivos. “Deus está nos detalhes.”

Outra formulação do fator de compreensão é a seguinte. Os implementadores não atuam como receptores passivos e objetivos. Um ator não é uma *tábula rasa*, *uma folha de papel em branco* que registra fielmente as mensagens que vêm de cima. As mensagens são

filtradas, classificadas e rotuladas por meio de *mapas cognitivos, esquemas de codificação, redes perceptivas, prismas ideais, paradigmas e outras estruturas mentais*. A recepção de sinais de governança de cima é estruturada por essas cognições de receptores. As pessoas veem aquilo que elas têm ideia do que seja, do contrário não entendem. É verdadeiro que receptores podem agir passivamente e, simplesmente aceitar o que lhe é requisitado; no entanto eles também podem ser ativos, seletivos, classificadores, transformadores e *estruturantes* (ver BOSWELL, 2009, p. 45; ver PATTON, 2011, p. 15 e 16). Isso, claro, afetará a implementação contínua e, finalmente, o *outcome* final.

Além da compreensão do implementador, *process tracing* deve levar em consideração a disposição, a boa vontade do implementador. A disposição pode dificultar ou promover o curso de ação indicado na intervenção (SABATIER; MAZMANIAN, 1980, p. 153 e 489; DOWNS, 1967; ELMORE, 1978, p. 191 et seq.; VAN HORN; VAN METER, 1977, p. 113). Nessa área, devem ser mencionados dois fragmentos principais da teoria: a teoria do comportamento burocrático autocentrado (*self-regarding*) e a teoria da incompatibilidade atitudinal.

As burocracias públicas podem entreter agendas próprias, o que pode entrar em conflito com a implementação fiel das diretrizes e das recomendações de seus diretores. Uma visão influente é a concepção de burocratas como maximizadores de orçamento. Para melhorar sua reputação pessoal, salários, campos de influência e outros objetivos egoístas, os burocratas tendem a expandir seus orçamentos. O seu ímpeto organizacional de crescer ou, pelo menos, de manter o *status quo* prevalece sobre a preocupação com a implementação fiel de políticas substantivas reais. Anulando-se como servidores públicos, os funcionários do Estado transformam-se em senhores sem espírito público. Esse pressuposto impeliu os pesquisadores da *tradição de escolha pública* a deduzir sobre o comportamento burocrático de maximização da utilidade (NISKANEN, 1971; LEWIN, 1988).

Os interesses egoístas são mais salientes, talvez, quando os partidos políticos e o público em geral querem cortar fundos do mecanismo governamental. Todos os corpos preocupados rotineiramente respondem reclamando em alto e bom som que cada corte de fundos e de pessoal reduzirá os serviços a níveis intoleráveis (JØRGENSEN, 1981; NIELSEN, 2010).

Uma variação mais branda do tema da *vontade* é que os servidores públicos civis podem ter dúvidas sobre a adequação do

conteúdo da política, o que pode dificultar sua supervisão. A *teoria da incompatibilidade atitudinal* indica que a ação de agência lenta ou descumpridora pode ser de dois tipos: baseada em fins e baseada em meios. Em caso de descumprimento baseado em fins, os funcionários responsáveis pela aplicação da lei desaprovam os objetivos da política (não conformidade) ou, pelo menos, atribuem a esses objetivos uma menor prioridade na competição por recursos escassos. O descumprimento baseado em meios, por sua vez, envolve dúvidas sobre a eficácia dos instrumentos recomendados de políticas públicas e os arranjos organizacionais para atingir os objetivos estabelecidos.

Nos Estados de bem-estar nórdicos, onde se supõem que os municípios implementem as políticas sociais do governo nacional, os líderes políticos nos comitês sociais municipais devem estar dispostos a obedecer e a cumprir as políticas sociais do Estado. Se o município é governado por partidos políticos de cores ideológicas diferentes daqueles que inauguraram as políticas em questão, isso pode influenciar a execução e vice-versa.

Capacidade, o terceiro fator da tríade de compreensão-vontade-capacidade, significa ter a aptidão de realizar algo, como, por exemplo, tomar um curso de ação indicado em uma intervenção. Se a execução de um programa requer dotações, pessoal qualificado, tempo, energia e equipamentos indisponíveis para a agência central, a probabilidade de alcançar resultados bem-sucedidos será enfraquecida. Uma situação comum é a *falta de recursos financeiros*. Com fundos insuficientes, as autoridades administrativas não podem contratar pessoal suficiente para processar pedidos de apoio econômico, monitorar efetivamente o cumprimento do marco regulatório ou conduzir análises técnicas envolvidas no desenvolvimento de novas regras e regulamentos.

Com esses comentários, deixa-se a tríade capacidade-compreensão-vontade, e volta-se brevemente à influência de indivíduos com características especiais.

Um programa pode ser implementado de forma eficiente devido aos esforços de *zelotes ou empreendedores*, ou seja, pessoas que estão fortemente dedicadas à sua tarefa e capazes de trabalhar duro com a implementação. Eles podem ser recém-formados que ainda não se encaixaram em nenhum molde administrativo ou substantivo em particular. Eles podem ser pessoas com experiência que trabalharam muito tempo no campo, que confiam em suas habilidades e se sentem muito seguras em seu papel profissional.

A troca de pessoas em uma entidade coletiva pode, portanto, ser importante. Se os *zelotes* desaparecerem para realizar outras atividades, os resultados desejáveis podem desaparecer; se *zelotes* entrarem na organização, a eficácia do programa pode aumentar de repente.

15.8 Implementação 3: Respostas do Operador de Nível de Rua

Os avaliadores de *process tracing* devem levar em consideração as atividades do operador de nível de rua porque podem afetar consideravelmente a execução e os resultados das políticas públicas⁷.

Os operadores de nível de rua são funcionários públicos que, em seu trabalho, estão em contato direto com membros do público em geral. Realizam as ações exigidas pelas políticas públicas em áreas que vão desde educação e serviços sociais até segurança, defesa e meio ambiente. Alguns exemplos incluem os inspetores de construção civil, policiais, guardas de fronteira, assistentes sociais e professores das escolas públicas. Esses funcionários públicos têm contato imediato com o público em geral, em contraste com os funcionários públicos que realizam análise de políticas ou análises econômicas, que não se encontram com o público. Os operadores de nível de rua implementam decisões políticas tomadas por altos funcionários do serviço público.

Os operadores de nível de rua interagem com o público em geral, pessoalmente, por telefone ou via internet⁸.

⁷ Na vida real, pode haver uma camada imediatamente acima dos operadores de nível de rua na cadeia de implementação. A expressão “burocrata de nível de tela” (BOVENS; ZOURIDIS, 2002) pode significar superintendentes imediatos de burocratas de linha de frente (SCHLAEGER, 2013, p. 74-77) (“panopticon virtual”); e acima daqueles, é claro, pode haver várias camadas mais.

⁸ Os sinônimos para operadores de nível de rua incluem funcionários de linha de frente, profissionais de linha, profissionais, agentes da lei, oficiais de campo e funcionários de classificação. Exemplos de estudos de operador de nível de rua: Kaufman (1967) (guardas florestais dos EUA); Stjernquist (1973) (guardas florestais, engenheiros florestais); Wilson (1978) (agentes de narcóticos do FBI); Wilson (1989, p. 27-89); Lundberg (1982) (inspetores do ambiente de trabalho); Municio (1987); Johansson (1992); Schierenbeck (2003) (burocratas da linha de frente, funcionários do trabalho na Suécia e Israel); Jewell (2007) (assistentes sociais); Høybye-Mortensen (2011); Durose (2011); Ellis (2011); Sevä (2015) (inspetores ambientais).

O famoso livro de Michael Lipsky, *Street-Level Bureaucracy*, tem sido particularmente influente. No mundo de Lipsky, os burocratas do nível da rua são “trabalhadores do serviço público que interagem diretamente com os cidadãos no decorrer de seus empregos e que têm um poder discricionário substancial na execução de seu trabalho” (1980, p. 3). São funcionários públicos que concedem acesso a programas governamentais e prestam serviços dentro deles. Professores, policiais, profissionais de serviço social são típicos burocratas do nível da rua do molde de Lipsky⁹.

Os burocratas de nível de rua criam políticas, argumenta Lipsky, por meio da multidão de decisões que tomam ao interagir com os clientes. Eles possuem poder discricionário porque nunca há recursos suficientes para fornecer supervisão próxima, frequente e direta, e, também, porque estão fisicamente separados dos seus superiores. Além disso, não existem critérios de desempenho precisos que especifiquem como os agentes prisionais, médicos em clínicas, engenheiros florestais, enfermeiros de saúde pública e assistentes sociais devem fazer seu trabalho. Em suma, mantém Lipsky, as políticas são formadas na implementação por operadores que desenvolvem rotinas e atalhos para lidar com os problemas no dia-a-dia dos seus públicos.

Uma vez que *a compreensão* e *a vontade* foram tratadas anteriormente, esta seção será focada no aspecto da *capacidade*. Novamente, um problema também para este segmento é a escassez de recursos.

Para algumas inspeções, a magnitude da dotação orçamentária legislativa torna praticamente impossível dispor de inspetores suficientes para observar os violadores. Os padrões de segurança e de saúde ocupacional aplicam-se a mais de 20 mil postos de trabalho; no entanto a *U.S. Board of Occupational Safety and Health (OSHA)* só pode inspecionar uma pequena fração deles todos os anos.

Restrições de tempo limitam a capacidade dos inspetores para descobrir infrações de normas e de regulamentos. Hemenway (1985, p. 11) enunciou que os inspetores da OSHA gastavam apenas cerca de um terço do tempo disponível no campo. O resto foi usado preparando-se para viajar e relatando sobre visitas de campo.

⁹ Pressman; Wildavsky (1984). Também Kelman (1981, p. 180 et seq.).

Os inspetores regionais da Comissão Regulatória Nuclear Norte-Americana (*American Nuclear Regulatory Commission*) gastavam apenas cerca de 25% de uma típica semana de trabalho na usina. Muito tempo valioso esteve dedicado ao trabalho clerical e mundano (também BARDACH; KAGAN, 1982, p. 123 et seq.).

Ao ajustarem-se à escassez de recursos, os operadores de programas adotam *estratégias de enfrentamento (coping strategies)*. Para evitar sobrecarga do número de casos, eles limitam a informação sobre seus serviços, pedem aos clientes que esperem na fila, se tornam indisponíveis para contatos ou usam amplamente as referências de clientes para outras autoridades. *Escolher seletivamente clientes* é outra saída. Os operadores da linha de frente concentram sua atenção em um número limitado de clientes selecionados, tipos de programas e soluções. Eles preferem casos fáceis e bem definidos do que os mais difíceis, amorfos e demorados. As estratégias de enfrentamento (*coping strategies*) são tão comuns que impedem a implementação em formas sistemáticas, o que impacta o *outcome* final, argumenta Lipsky (1980).

15.9 Implementação 4: Respostas do Grupo-Alvo

O quarto determinante-chave do núcleo da implementação que avaliadores que usam *process tracing* devem ter em mente diz respeito à resposta do grupo-alvo. Na tríade *compreensão-capacidade-vontade*, apenas algumas observações sobre capacidade e sobre vontade podem ser suficientes como ilustrações.

O tempo de adaptação da intervenção à situação de decisão do público-alvo é um fator de *capacidade* a ser considerado no processo de rastreamento. Se chegar em um *momento formativo*, a intervenção pode produzir uma resposta positiva imediata; caso contrário, não. Suponha que o Estado esteja disseminando informações para empresas industriais, sugerindo que elas deveriam economizar eletricidade por meio da instalação e do uso racional de equipamentos de última geração em termos de conservação de energia. Se a mensagem chegar quando a empresa estiver a ponto de renovar a maquinaria, ela poderá ser efetuada imediatamente. Em contraste, se a empresa acabou de mudar seu equipamento, o *momento formativo* dissipou-se. A informação cairá em terreno infértil, e as recomendações permanecerão desatendidas. A concepção temporal do alvo não atingiu a do governo¹⁰.

O Estado pode tentar alterar os padrões de comportamento do público-alvo, sendo que parte dos destinatários estão contra ou têm pouco interesse em aprovar mudanças. Os requisitos governamentais de equipamentos com eficiência energética ou um bom ambiente de trabalho podem ser exemplos disso, pois podem entrar em conflito com os desejos das empresas para mostrar boa produtividade e grandes lucros.

Os regulados, os supostos beneficiários e outros tipos de destinatários podem, obviamente, cumprir corretamente as prescrições e as proibições regulatórias. No entanto, para evitar a conformidade, os grupos-alvo podem tentar *capturar* funcionários da agência, ou seja, ganhar controle e exercer influência sobre eles. Embora a captura possa ocorrer em todos os níveis de implementação, ela será aplicada apenas aos encontros entre grupos-alvo e inspetores individuais (HEMENWAY, 1985).

A inspeção regulatória é intrinsecamente adversa. Normalmente, a tarefa do inspetor é descobrir violações de prescrições governamentais. O inspecionado, por sua vez, geralmente prefere que nenhuma infração seja encontrada. Em situações como esta, inspetor e inspecionado têm objetivos conflitantes.

No entanto, o relacionamento não é de puro conflito. Em muitas facetas de sua interação, a cooperação pode melhorar. Ambos querem ver alguns problemas corrigidos. Ninguém quer uma catástrofe nuclear, um acidente de avião ou uma intoxicação alimentar. (HEMENWAY, 1985, p. 13).

Os principais beneficiários das inspeções são, frequentemente, interesses dispersos, não organizadores – usuários de elevadores, pedestres, os que gostam de comer maçãs, nadadores, residentes das casas de repouso, etc. Muitos deles nunca veem o inspetor, nem mesmo entendem seu trabalho. Raramente o recompensarão por executar um bom trabalho ou o castigarão por fazê-lo mal.

Em contrapartida, o inspetor conhece quem é inspecionado. Muitas vezes, a interação está em andamento e face a face. O inspetor tem o poder de impor grandes custos ao inspecionado, o qual tem, desse modo, o motivo e a oportunidade de influenciar o comportamento do inspetor.

¹⁰ Outras noções na literatura para aproximadamente o mesmo incluem “pontuação” na teoria do equilíbrio pontuado, “Kairos” (momento frutífero), “janela de oportunidade”, “janela de política” e “junção crítica” ou “conjuntura crítica”.

O inspetor está mais propenso à cooptação quando os exames são longos, face a face e repetidos. Ele é mais facilmente capturado ao monitorar apenas uma indústria ou uma empresa. Hemenway (1985, p. 52) cita um inspetor de carnes e aves:

[Nós] saímos para inspecionar os regulados para fazer o nosso trabalho. Nós não só os visitamos periodicamente, como nós praticamente nos casamos com eles. Dia após dia, noite após noite, estamos na toca do leão sozinhos com o leão. Como será que nos damos? O Departamento de Agricultura dos EUA (*United States Department of Agriculture- USDA*) não nos conta. Como devemos resistir à barragem de ameaças e de tentações que os empacotadores de carne e frigoríficos constantemente nos dirigem? O USDA não nos conta. O USDA nos diz para usar nossa engenhosidade para fazer o nosso trabalho, para usar o nosso senso comum - mas isso não é muito útil quando você está na toca do leão.

O inspetor pode ser influenciado por incentivos psíquicos ou recompensas materiais. Os subornos são o caso paradigmático, mas métodos mais sofisticados podem ser usados. Perspectivas de futuros empregos bem remunerados são uma recompensa para alguns. Se os inspetores planejam buscar emprego com a indústria regulamentada, cabe-lhes manter uma relação amigável e não contraditória. Tudo isso é propício para a captura.

Às vezes, os inspecionados podem *camuflar* suas infrações. O uso exitoso dessa abordagem de dissimulação de violações depende, em grande parte, da natureza dos itens inspecionados. Alguns problemas são mais difíceis de encobrir do que outros. É mais fácil esconder violações de conduta do que falhas com equipamentos de capital. Hemenway (1985, p. 15) argumenta:

É fácil para o bombeiro garantir que o sistema de *sprinklers* esteja instalado corretamente ou para o inspetor de habitação determinar se o *boiler* ou a pia funcionam (ou mesmo se o banheiro tem uma pia) ou para que o inspetor

de restaurantes observar se o termômetro na câmara frigorífica fornece leituras precisas. É mais difícil monitorar o comportamento. É difícil para o bombeiro saber se as portas de saída ficarão ocasionalmente obstruídas, ou para o inspetor de habitação notar se o lixo às vezes é deixado nos corredores, ou para o inspetor de restaurantes garantir que os funcionários sempre lavam as mãos, coloquem seus cabelos longos em uma rede, e mantêm o abridor de latas livre da sujeira.

Uma possibilidade diferente é que os destinatários, por sua própria iniciativa e despesas, tenham planejado implementar as medidas em questão, mas participem do programa de qualquer forma, por exemplo, recebendo apoio econômico. Os destinatários são “caronas” (*free riders*) no programa do governo. As consequências são “efeitos de peso morto”, ou seja, impactos que teriam sido produzidos também sem a intervenção do governo. A *adicionalidade* zero indica o mesmo que isto: a quantidade de resultado com o suporte ao programa em comparação ao *outcome* sem o suporte é nula.

Para obter efeitos programáticos reais, os *process tracers* devem notar as consequências do “carona” (*free rider*) e subtraí-los dos efeitos gerais. (SANDBERG; FAUGERT, 2007, p. 23 et seq.). Obviamente, as reações dos destinatários são cruciais para a eficácia do *outcome* da intervenção.

15.10 Três Contextos que cercam o núcleo da implementação

Os impactos contextuais são importantes e devem ser levados em consideração em *process tracing*. *Process tracing* parte do pressuposto de que as intervenções disponibilizam recursos aos destinatários, cujas interpretações podem afetar suas considerações, raciocínio e ações concomitantes, dependendo do contexto em que tudo isso ocorre. A palavra “contexto” refere-se ao conjunto de circunstâncias (cercanias, ambientes) em que ocorre a intervenção avaliada e o curso de sua implementação básica.

O contexto é dividido em três tipos. O “Contexto 1: a formação da intervenção e a sombra do futuro” abrange atores, atividades e eventos no passado, quando a intervenção foi concebida, deliberada,

desenvolvida e sugerida para adoção. Contém elementos que podem influenciar a implementação tanto posteriormente, como antes da adoção formal da intervenção.

O segundo tipo, “Contexto 2: intervenções paralelas, outros *stakeholders*, cidadãos, mercados”, é mais extenso, variado e complexo. Ele abrange os impactos das circunstâncias contemporâneas, ou seja, contingências no ambiente que envolve a implementação do núcleo enquanto ela se desenvolve. Estão incluídas outras intervenções públicas paralelas à que está focada; indivíduos comprometidos que denunciam violações de leis e códigos morais em seus negócios e organizações (denunciantes); cidadãos por desobediência civil; e mercados através de mudanças nas demandas dos consumidores e flutuações de preços.

O terceiro e último tipo, “Contexto 3: Monitoramento e avaliação *ex-post*”, é composto por avaliações pós-adoção da intervenção e sua implementação.

15.11 Contexto 1: Formação da Intervenção e Sombra do Futuro

Primeiramente, tratar-se-á da “formação da intervenção”. A busca pela origem e pela história formativa de intervenções em curso ou encerradas pode parecer supérflua quando a finalidade do *process tracing* é avaliar sua implementação e seus impactos. Por que olhar para trás quando o impulso da sua atividade avança? É verdade que *process tracing* se concentra intensamente nos processos entre a intervenção adotada, por um lado, e seus possíveis resultados posteriores, por outro. No entanto, as forças que operam anteriormente na fase de formação de intervenção podem muito bem influenciar os processos de implementação mais tarde. É isso que se chama de sombra do futuro.

A característica da participação pode ilustrar. A *participação* em processos de intervenção anteriores pode muito bem influenciar os resultados de intervenção posteriores. A intervenção inaugurada pode ter sido precedida por um ou por vários inquéritos especiais de investigação com muitos *stakeholders* envolvidos, conflitos partidários, atenção da mídia e procedimentos formais de tomada de decisão nos escritórios do governo central. Se os potenciais *stakeholders* afetados tiverem permissão para contribuir com ideias e com argumentos durante a formação passada, a reforma resultante será infundida

com uma legitimidade, que se transformará em apoio moral e esforços positivos ativos durante sua implementação futura. Um programa formado por meio de um processo, onde a autoridade pública, que mais tarde será encarregada da sua implementação, terá permissão para participar, provavelmente será implementada mais pronta e fielmente do que as intervenções desenvolvidas e adotadas acima do chefe da autoridade pertinente. O caso extremo é quando a própria agência iniciou a reforma e pressionou sua adoção. Evidentemente, uma vez inaugurada, tais reformas serão mais facilmente realizadas.

O mesmo raciocínio aplica-se à participação de partes afetadas além da agência nacional, como órgãos de implementação regionais e locais, intermediários privados e clientes. Se, por exemplo, as organizações de clientes pressionam fortemente por uma intervenção e participaram do seu desenvolvimento passado, será mais aceitável para eles do que se eles não tivessem participado na sua formação. Isso provavelmente aumentará a legitimidade, facilitará o cumprimento e diminuirá a potencial oposição.

A implementação rápida e fiel pode vir de duas maneiras. Primeiro, as características dos processos que se desdobram *antes* podem influenciar a implementação de uma intervenção, uma vez que somente algum tempo *depois* é inaugurada. O mecanismo de conexão é a *memória*, ou seja, a *habilidade humana de lembrar*. A implementação pode começar imediatamente após a adoção formal da intervenção e com determinação devido a boas lembranças de como a intervenção se originou. A implementação também pode começar tarde e, de má vontade, devido às más lembranças da fase de formação da intervenção. Sua participação anterior desenvolveu neles (implementadores) sentimentos de responsabilidade e de legitimidade (ou irresponsabilidade e ilegitimidade) em relação ao produto resultante e, conseqüentemente, à sua implementação.

Em segundo lugar, os traços nos processos que se desdobram antes podem influenciar a implementação de uma intervenção tecnicamente não existente uma vez que se desenrolam. O mecanismo é a antecipação humana. A implementação pode começar antes da adoção formal da intervenção porque a probabilidade percebida de que uma intervenção sugerida será formalmente adotada posteriormente é tão alta que os atores executivos começam sua implementação, embora ainda não seja formalmente autorizado.

Na ordem política sueca, o trabalho de implementação já pode começar quando alguma comissão de investigação *ad hoc* do governo, após meses de inquérito, apresentar oficialmente seu relatório final escrito com propostas que exijam ação governamental. Em alguns casos, a implementação pode começar uma vez que tal comissão é nomeada e tenha seu mandato conferido pelo governo, ou seja, antes de iniciar qualquer pesquisa de informações. Às vezes, o mecanismo da política pública é tão previsível que os atores interessados podem concordar ou contornar uma próxima intervenção antes da sua adoção formal. Eles atuam *na sombra do futuro* (“*na sombra da intervenção*”).

Após o embargo de petróleo árabe de 1973, o governo sueco repetidamente usou o tabelamento de preços como um instrumento de política para reduzir os preços. Estudos depois mostraram que tanto os empregadores quanto os funcionários aprenderam a prever o próximo tabelamento de preços antes de serem inaugurados e apressaram-se a aumentar seus preços e salários para já terem sido previamente pagos quando o tabelamento finalmente chegasse. Este mecanismo de antecipação impulsionou preços e salários em vez de detê-los - um efeito perverso não pretendido e indesejado. (JONUNG, 1984, p. 208).

Em fevereiro de 1991, o boletim “Kommun Aktuell” (“Município Atual”) apresentou uma outra história de antecipação (*anticipation story*): “Um terço dos municípios já possui propostas completas para uma nova organização do conselho. Apesar de o governo não ter apresentado sua proposta de um novo Acto Municipal [e ainda mais notável: sem a aceitação pelo Parlamento da proposta do governo].” Os municípios anteciparam a decisão futura do Estado e cumpriram-na bem antes de serem obrigado a fazer isso. Eles agiram na sombra do futuro (VEDUNG; SVÄRD, 2008).

Aqueles que ignoram o fenômeno da atuação na sombra do futuro subestimam grosseiramente as influências das intervenções públicas.

Agora, será deixado o contexto da formação histórica - “a paisagem a partir da qual a intervenção decola” - e será dada atenção ao ambiente contemporâneo da intervenção.

15.12 Contexto 2: Intervenções Paralelas, Outros Stakeholders, Cidadãos, Mercados

Os contextos são de enorme importância para os avaliadores que usam *process tracing*. Um objeto facilmente implementado em um contexto pode encontrar barreiras difíceis em outro. A seção anterior foi dedicada aos impactos de fatores históricos na implementação posterior. Esta seção será dedicada aos impactos dos *determinantes do contexto contemporâneo*. É uma categoria ampla e complicada. São incluídas intervenções públicas paralelas, funcionários individuais comprometidos agindo como denunciante, cidadãos através de desobediência civil e mercados por meio de mudanças nas demandas dos consumidores e nas mudanças de preços.

Primeiro, considere *intervenções públicas paralelas* próximas. Os estudiosos de metodologia de avaliação advertiram os avaliadores contra a concentração de suas avaliações em uma determinada intervenção por vez. É importante especificar esse aviso. É perfeitamente correto avaliar uma única intervenção no sentido de verificar seu valor na produção de *outcomes* desejáveis e indesejáveis. No entanto, na tentativa de rastrear os *outcomes* esperados e todos os tipos de *spillovers*, é necessário adotar uma abordagem sistêmica para pintar uma imagem conformativa de um conjunto de fatores que, ao lado desta única intervenção, podem influenciar os *outcomes*. Por exemplo, outros programas paralelos no *mesmo setor* podem fortalecer ou enfraquecer a força implementadora da intervenção. Da mesma forma, políticas paralelas em *outros setores* (jurisdições) podem influenciar *outcomes* diretamente, ou indiretamente, por meio da implementação da intervenção. Desse modo, os resultados podem ser influenciados por agências e por redes de atores trabalhando com intervenções diferentes, embora paralelas, em nível nacional, regional e local. (HJERN, 1982; WINTER, 1990; ELMORE, 1980; HANF, 1982 e HANF; TOONEN 1985).

A *teoria do conglomerado da administração pública* (*conglomerate theory of public administration*) é apresentada aqui na tentativa de capturar o caráter sistêmico e configurador da situação. Um conglomerado (C1, C2) é uma coisa (uma entidade) que consiste em várias partes ou itens que são agrupados. A literatura descreve fenômenos semelhantes como "*coalizões de defesa*

(*advocacy coalitions*)”, “*triângulos de ferro*”, “*redes de questões*”, “*redes de políticas*”, “*comunidades de políticas públicas*” e “*estruturas de implementação*” (HJERN; LUNDMARK, 1979; LUNDQVIST, 1987, p. 66 et seq.).

As políticas nacionais suecas para o aproveitamento de rios indômitos para geração de eletricidade podem ilustrar a teoria do conglomerado. Por intermédio de seu Conselho Estatal de Energia, o Estado sueco quer transformar rios selvagens em fontes de energia hidrelétrica, construindo represas de reservatório, usinas de água, túneis de águas subterrâneas e canais de desvio de água de uma área de captação para outra para geração de eletricidade. Ao longo do tempo, o governo também procura elaborar políticas regionais para distribuir a população e as políticas do mercado de trabalho para combater o desemprego. Como a maioria dos rios indômitos estão nas regiões norte e ocidental do país, as políticas regionais e do mercado de trabalho apoiam a sua exploração.

Simultaneamente, o governo nacional quer salvar alguns cursos de água por razões ambientais, cênicas e culturais. Dessa forma, os conselhos nacionais para o meio ambiente e os assuntos culturais tendem a desaprovar a exploração. Uma vez que as barragens vão submergir as florestas estatais, a agência encarregada da silvicultura estatal e operada pelo governo pode se contrapor ao Conselho Estatal de Energia. A exploração do potencial energético da água pode prejudicar a indústria agropecuária das renas (*reindeer industry*) dos lapões (*Laplanders* – o povo *Sami*, moradores da Lapland (Lapônia), os Sápmi no extremo norte da Suécia, Noruega, Finlândia and the Kola peninsula in Russia), o que induzirá o Colegiado de Câmara (*Chamber Collegium*), que gerencia as políticas da Lapônia, a se mobilizar contra uma maior exploração hidráulica.

Em resumo, longe de se unir atrás de um objetivo comum, diversos atores do setor público apresentam requisitos contraditórios para os rios indômitos. Isso torna a implementação de políticas de exploração de energia hidrelétrica e outras políticas também controversas e passíveis de entrar em conflitos.

O apoio ou a resistência de *stakeholders não governamentais* (“outros *stakeholders*”) podem influenciar os resultados. Organizações e companhias com fins lucrativos ou sem fins lucrativos que tenham sido afetadas podem encorajar ou desencorajar agências falando e atuando de várias maneiras.

Pegue a Lei Sueca de Exportação de Armas, por exemplo, a qual proibiu a venda de equipamentos militares fabricados na Suécia para países em guerra. No entanto, não era a Inspeção de Materiais de Guerra, o supervisor oficial da Lei, que revelou que a indústria de armas sueca, em particular Bofors, vendeu armas ao Irã durante sua guerra com o Iraque, usando pessoas em Cingapura como intermediários. As violações do ato foram descobertas pela Associação de Paz e Arbitragem Sueca, que é baseada na atuação cidadã. Uma organização não governamental sem fins lucrativos, sem papel formal na administração do ato, passou a influenciar o *outcome*.

Um caso especial é *denunciar*: o relatório de um funcionário sobre a violação da lei pelo empregador. Nos regulamentos de monitoramento, os inspetores podem obter ajuda dos funcionários anônimos descontentes que oferecem dicas e expõem as infrações da empresa. Isso foi o que aconteceu no caso Bofors mencionado acima.

Ou considere a *desobediência civil*. "A desobediência civil é a recusa ativa e professada de obedecer certas leis, demandas ou ordens de um governo ou de um poder internacional de ocupação" (WIKIPEDIA; CIV, 2017). Os *process tracers* devem procurar a desobediência para dificultar ou para acelerar a implementação, realizada publicamente ou, pelo menos, publicamente anunciada.

Exemplos escandinavos de desobediência civil incluem a implementação local de decisões de alto nível para construir estradas através de florestas cênicas e parques municipais, ou para construir usinas de energia em rios selvagens. Os protestos contra a exploração dos rios Alta e Mardøla na Noruega e o corte planejado de árvores olmos do Jardim do Rei na cidade de Estocolmo (*Kungsträdgården*) são exemplos óbvios de que a implementação legal pode enfrentar a resistência do cidadão.

Um fator que, às vezes, não é lembrado em *process tracing* é o poder do consumidor nos *mercados*. Aqui, a palavra "consumidor" não se refere aos objetivos pretendidos ou destinatários de processos de implementação comuns; "consumidores" nesse sentido foram discutidos em "Implementação 4: resposta do grupo-alvo" acima. Aqui, refere-se aos consumidores de produtos da empresa, uma vez que as empresas - mas não seus consumidores - constituem os alvos de uma intervenção pública, como um regulamento, por exemplo.

As respostas negativas ou a rejeição dos consumidores a bens problemáticos podem obrigar os produtores regulados a cumprir com

mais prontidão e fidelidade do que de outra forma. Reações do consumidor podem ser reportadas nos meios de comunicação de massa ou notado em estatísticas de vendas combinadas com pesquisas de boca a boca ou de consumidores conduzidas pelas próprias empresas. Os produtores também podem antecipar as reações negativas dos consumidores ainda não ocorridas e agir proativamente, a fim de evitar debates críticos na mídia prejudicando a imagem e a marca registrada da empresa e, a longo prazo, prejudicando futuras vendas.

Se uma intervenção pública tentar reduzir o uso de um produto químico perigoso, por exemplo, os consumidores podem parar de comprar produtos que contenham esse produto químico e, assim, contribuir para a sua eficácia. Electrolux, o grande produtor sueco de aparelhos de refrigeração, terminou a produção de refrigeradores com substâncias CFC (clorofluorocaboreto) como líquido refrigerante antes que uma regulamentação governamental determinasse que eles tivessem que o fazer devido aos consumidores alemães e holandeses que exigiam apenas refrigeradores *freon-free*. O CFC em refrigeradores foi considerado a maior ameaça, e o grande responsável pela destruição da camada de ozônio estratosférica protetora da Terra. (VEDUNG; KLEFBOM, 2002, p. 203 e 204, p. 314 et seq.).

Uma vez que os *preços de mercado* aumentam, a implementação das intervenções é facilitada ou obstruída. Na Suécia, a principal razão para o declínio do consumo de energia desde o início da década de 1970 até a década de 1980 foi, provavelmente, o aumento substancial dos preços do petróleo entre 1973 e 1979. Esses aumentos de preços foram produzidos no sistema internacional. As políticas de conservação de energia do governo sueco precipitaram e facilitaram o desenvolvimento; mas, em grande parte, não o produziram.

15.13 Contexto 3: Monitoramento e Avaliação Ex-Post

Um terceiro e último amplo grupo de fatores de contexto para tomar nota em *process tracing* para avaliação está enraizado no monitoramento e na avaliação *ex-post* - ou seja, quando a intervenção for formalmente adotada. A administração, o *output* e o *outcome* podem ser facilitados (ou frustrados, dificultados) porque os decisores de intervenção persistentemente e implacavelmente endossam (ou falham em endossar) a intervenção uma vez que ela é inaugurada.

As demonstrações de endosso incluem declarações reiteradas sustentando uma interpretação particular da intervenção, ou declarações repetidas sobre a importância da execução fiel e rápida. Essas mensagens podem ser permitidas pelos implementadores para influenciar suas atividades e afetar os *outcomes*. As mensagens insistentes e saturantes repetidas pelos diretores podem criar pressões de opinião que conduzam a sérios esforços de implementação por agências, intermediários e burocratas de linha de frente e, dessa forma, influenciam *outcomes*, *ceteris paribus*. Trata-se de *monitoramento informal* (aferição informal), uma vez que não se baseia em monitoramento/avaliação investigativa cuidadosa.

O monitoramento informal pode ser formulado como *reivindicação de crédito* (PIERSON, 1994). Reivindicar o crédito significa “anunciar-se como responsável por realizar algo (geralmente bom)” (*Credit claiming 1*). Na política pública, a reivindicação de crédito é definida como “agir de modo a gerar uma crença em um ator político relevante (ou atores) que é pessoalmente responsável por fazer com que o governo, ou alguma unidade dela, faça algo que o ator (atores) considera desejável” (*Credit claiming 2*). A reivindicação de crédito ocorre quando um instigador de intervenção A por declarações e ações tenta criar uma imagem nos implementadores B, C e D que A, o partido político de A, a coalizão do governo de A ou a maioria de A é responsável por uma intervenção adotada é *substancialmente* necessária ou extremamente urgente para o país ou para um importante grupo de destinatários, e que, portanto, ele deve ser executado com firmeza e destreza. A motivação para isso pode ser que a invenção seja substantiva ou ideologicamente boa, ou seja, de acordo com o que se acredita ser bom para a sociedade.

No entanto, há também a *motivação estratégica*. Desde que a política seja popular, o argumento do reivindicador de crédito pode ser criar uma imagem favorável de seu partido no eleitorado, em vários grupos de *stakeholders*, nos meios de comunicação de massa, etc., para conquistar votos nas próximas eleições.

A aferição informal também aparece como *evitação (esquivar-se) da culpa*. Aos olhos dos instigadores, a política adotada pode ter parecido substancialmente promissora; mas, mais tarde, ela desenvolve-se em uma direção tão negativa que eles não querem se

associar a ela. Quando a política é atacada na mídia e o público em geral expressa opiniões negativas sobre isso, os diretores abster-se-ão inteiramente de articular o apoio à política ou falarão muito pouco e em voz baixa sobre ela. Além de considerações substantivas, provavelmente também há razões estratégicas por detrás das posturas dos diretores.

No entanto, o silêncio é interpretado negativamente pelos implementadores. Isso torna-os hesitantes, irresolutos, incrédulos e desanimados em seu trabalho de implementação. O silêncio, quando os implementadores esperam e anseiam por apoio, afeta negativamente a implementação e os resultados.

Se a aferição informal for um subtipo do Contexto 3, a *aferição formal* também é na forma de *monitoramento/avaliação* plena. O monitoramento formal pós-adoção oferece oportunidades para que os envolvidos supervisionem se a implementação se desenrola conforme desejado e esperado. Caso contrário, o monitoramento de dados indica que algo deve ser feito. Dessa forma, o monitoramento pode influenciar a implementação.

O monitoramento envolve a coleta de dados e a elaboração de estatísticas sem relacionar a informação coletada com as expectativas de intervenção ou com os critérios de valor expressos. O monitoramento é visto como um exercício estatístico não tendencioso, e ponto final. No entanto, essa informação quantitativa simples permite aos atores afetados verificar a forma como a implementação evolui e tomar as medidas adequadas.

Mas o “monitoramento” também pode constituir um tipo de avaliação. O acompanhamento nesse segundo sentido - avaliação formativa, avaliação do processo ou acompanhamento qualificado - não é apenas sobre a coleta de dados sobre o estado de coisas em determinadas variáveis. Envolve a resolução de problemas e o trabalho de reparo nos estágios de saída e pré-saída da implementação. O acompanhamento qualificado pode acompanhar todo o processo administrativo até a equipe de campo operacional e seu resultado final, a fim de esclarecer a situação em relação ao solicitado, esperado, desejado ou indesejado. O monitoramento qualificado direciona seu foco de pesquisa de acordo com a conformidade e os desvios na implementação.

Mas não só isso. O monitoramento também procura identificar causas motoras por trás das conformidades e desvios e formulá-los como explicações. Também busca por fatores que podem estar por trás de situações surpreendentemente ruins ou boas. Em outras palavras, trata-se de análises causais sob a forma de busca de determinantes (ver VEDUNG, 2013, p. 58 e 59 e VEDUNG; DAHLBERG, 2013, p. 104 e análise a montante).

Por que os avaliadores que empregam *process tracing* devem verificar se a análise de desvios e a busca de causas nos acompanhamentos e avaliações qualificados realizados funcionam em relação ao trabalho de implementação na intervenção em estudo? Exatamente porque a finalidade dessas avaliações formais é melhorar o desempenho do programa, então a tarefa de *process tracing* é descobrir se e em que maneiras eles (acompanhamentos/avaliações qualificadas) interferiram nos processos de implementação. Por acaso, as avaliações formais encontraram quaisquer barreiras de implementação e foram estas eliminadas graças às avaliações? Essa poderia ser uma das tarefas principais dos avaliadores de *process tracing* nessa etapa.

Isso encerra a pesquisa configurativa de fatores muito amplos e abrangentes, o que pode contribuir para explicar os *outcomes* das intervenções públicas. Espera-se que este exercício forneça orientação na execução de avaliações de *process tracing*.

REFERÊNCIAS

ALEXANDERSSON, Karin. **Vilja, kunna, förstå: Om implementering av systematisk dokumentation för verksamhetsutveckling i socialtjänsten** [Willingness, comprehension, capability: On implementation of systematic documentation for activity development in public social work]. Örebro: Örebro University Library, PhD thesis, 2006. English summary, www.diva-portal.org/oru/abstract.xsql?dbid=449 (accessed 2018072)

ALLISON, Graham T. Implementation Analysis: The 'Missing Chapter' in Conventional Analysis Illustrated by a Teaching Exercise. In LEWIN, Leif. VEDUNG, Evert (Eds.). **Politics as Rational Action: Essays in Public Choice and Policy Analysis**. Dordrecht, Holland: D. Reidel Publishing Co., 1980, p. 237 260.

ANDERSON, Scarvia B.; BALL, Samuel. **The Profession and Practice of Program Evaluation**. San Francisco: Jossey-Bass, 1978.

ANDERSON, Scarvia B.; BALL, Samuel; MURPHY, Richard T. (Eds.). **Encyclopedia of Educational Evaluation**. San Francisco: Jossey-Bass, 1974.

ANONYMOUS PEER REVIEW, 2013, http://en.wikipedia.org/wiki/Anonymous_peer_review#Anonymous_peer_review (retrieved 20131205).

ARGYRIS, Chris. **Reasoning, Learning, and Action: Individual and Organizational**. San Francisco: Jossey-Bass, 1982.

ÅSTRAND, Kerstin. **Energy Policy Instruments: Perspectives on Their Choice, Combination and Evaluation**, Lund: Lunds universitet. Miljö- och energisystemstudier, 2006.

BALDWIN, David A. **Economic Statecraft**. Princeton, N.J.: Princeton University Press, 1985.

BARDACH, Eugene. **The Implementation Game: What Happens After a Bill Becomes a Law**. Cambridge, MA: The MIT Press, 1977.

BARDACH, Eugene; KAGAN, Robert A. **Going by the Book: The Problem of Regulatory Unreasonableness**. Philadelphia: Temple University Press, 1982.

BARRETT, Susan; FUDGE, Colin. (Eds.). **Policy and Action: Essays on the Implementation of Public Policy**. London: Methuen, 1981.

BEACH, Derek; PEDERSEN, Rasmus Brun. **Process-Tracing Methods: Foundations and Guidelines**. Ann Arbor, Michigan: The University of Michigan Press, 2013.

BEMELMANS-VIDEC, Marie-Louise; RIST, Ray C.; VEDUNG, Evert. (Eds.). **Carrots, Sticks, and Sermons: Policy instruments and Their Evaluation**. New Brunswick, NJ: Transaction, 1998

BEMELMANS-VIDEC, Marie-Louise; VEDUNG, Evert. Conclusions: Policy Instruments types, Packages, Choices, and Evaluation. In: BEMELMANS-VIDEC, Marie-Louise; RIST, Ray C.; VEDUNG, Evert. (Eds). **Carrots, Sticks, and Sermons: Policy Instruments and Their Evaluation**. New Brunswick, NJ: Transaction, 1998, p. 249-274.

BENGTSSON, Bo. Tenant's Dilemma – On Collective Action in Housing. **Housing Studies**, Vol. 13 (1998), p. 99-120.

BENGTSSON, Bo; ANNANIASSEN, E; JENSEN, L., RUONAVAARA, H.; SVEINSSON, J. R. (Eds.). **Varför så olika? Nordisk bostadspolitik i jämförande historiskt ljus** [Why So Different? Nordic Housing Policy in Comparative Historical Light]. Malmö: Egalité, 2006.

BENNETT, Alexander L. Process Tracing and Causal Inference. In BRADY, Henry E.; COLLIER, David (Eds.) **Rethinking Social Inquiry**. Lanham, MD: Rowman & Littlefield, 2010 (2nd edition), p. 207-19.

BERK, Richard A; BORUCH; Robert F; CHAMBERS, David L; ROSSI, Peter H; WITTE Ann D. Social Policy Experimentation: A Position Paper. **Evaluation Review**, Vol. 9 (1985), p. 387-440.

BERMAN, Paul. The Study of Macro- and Micro-implementation. **Public Policy** Vol. 26, n. 2 (1978), p. 157-184.

BERMAN, Paul. Thinking About Programmed and Adaptive Implementation: Matching Strategies to Situations. In INGRAM, Helen; MANN, Dean (Eds.). **Why Policies Succeed or Fail**. Beverly Hills: Sage Publications, 1980, p. 205-227.

BERNSTEIN, Ilene N.; FREEMAN, Howard E. **Academic and Entrepreneurial Research: The-Consequences of Diversity in Federal Evaluation Studies**. New York: Russell Sage Foundation, 1975, p. 187.

BICKMAN, Leonard. (Ed.). **Advances in Program Theory**. San Francisco: Jossey-Bass, 1990.

BOSWELL, C. **The Political Uses of Expert Knowledge: Immigration Policy and Social Research**. Cambridge: Cambridge University Press, 2009.

BOUDON, Raymond. **The Unintended Consequences of Social Action**. London: Macmillan, 1982.

BOVENS, Mark; ZOURIDIS, Stavros. From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control. **Public Administration Review** Vol. 62, n. 2 (2002), p. 174-184.

BREWER, John; HUNTER, Albert. **Multimethod Research: a Synthesis of Styles**. Newbury Park, CA: Sage, 1989.

BREYER, Stephen. **Regulation and Its Reform**. Cambridge, Mass: Harvard University Press, 2009.

BRYMAN, Alan. **Social Research Methods**, Oxford: Oxford University Press, 2004 (2nd edition).

BRYMAN, Alan. **Social Research Methods**, Oxford: Oxford University Press, 2008 (3rd edition).

CAMPBELL, Donald T. Considering the Case against Experimental Evaluations of Social Innovations. **Administrative Science Quarterly**, Vol. 15, n. 1 (1970), p. 110-113.

CAMPBELL, Donald T. Reforms as Experiments. **American Psychologist**, Vol. 24, n. 4 (1969), p. 409-429. Republicado com revisões em CAPORASO, James A.; ROOS, Leslie L. (Eds.). **Quasi-experimental Approaches: Testing Theory and Evaluating Policy**, Evanston, IL: Northwestern University Press, 1973, p. 167-225.

CAMPBELL, Donald T. Experiments as Arguments. **Knowledge: Creation, Diffusion, Utilization**, Vol. 3, n.3 (1982), p. 327-337.

CAMPBELL, Donald T.; STANLEY, Julian C. **Experimental and Quasi-experimental Designs for Research**. Chicago: Rand McNally, 1966.

CAMPBELL, Donald T.; STANLEY, Julian C. **Experimental and Quasi-experimental Designs for Research**. Chicago: Rand McNally College Publishing Company, 1973.

CAPORASO, James A.; ROOS, Leslie L. (Eds.). **Quasi-experimental Approaches: Testing Theory and Evaluating Policy**. Evanston, Ill: Northwestern University Press, 1973.

CARO, Francis. **Readings in Evaluation Research**. Russell Sage Foundation, 1971.

CC1 <http://apgovfall08.wikifoundry.com/thread/1863359/Define+and+give+-examples+of+credit+claiming> (accessed 20170902)

CHELIMSKY, Eleanor. (Ed.). **Program Evaluation: Patterns and Directions**. Washington, D-C: American Society for Public Administration, 1985.

CHEN, Huey-Tsyh. **Theory-Driven Evaluations**. Newbury Park, CA: Sage, 1990.

COLEMAN, James S. **Models of Change and Response Uncertainty: An Introduction to Mathematical Sociology**. New York: Free Press, 1980

COLLIER, David; LEVITSKY, Steven. Conceptual Hierarchies in Comparative Research. In COLLIER, David; GERRING, John (Eds.) **Concepts and Method in Social Science: The Tradition of Giovanni Sartori**. London: Routledge, 2009, p. 269-288.

COLLIER, David. Understanding *Process Tracing*. **PS: Political Science and Politics**, Vol. 44, n. 3 (2011), p. 823-830.

COLLIER, John Jr.; COLLIER, Malcolm. **Visual Anthropology: Photography as a Research Method**. Albuquerque, NM: University of New Mexico Press, 1986.

COOK, Thomas D.; CAMPBELL, Donald T. **Quasi-experimentation: Design and Analysis for Field Setting**. Boston: Houghton Mifflin, 1979.

CRESSWELL, John W. **Qualitative Inquiry and Research Design: Choosing among Five Approaches**. Sage Publications, 2013 (3rd edition).

CRONBACH, Lee J. et al. **Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements**. San Francisco: Jossey-Bass, 1980.

DAHL, Robert A. **Democracy and Its Critics**. New Haven, CT: Yale University Press, 1989. (Tradução em português: DAHL, Robert A. **A democracia e seus críticos**. Tradução Patrícia de Freitas Ribeiro; revisão da tradução Aníbal Mari. - São Paulo: Editora WMF Martins Fontes, 2012. - (Biblioteca Jurídica WMF)

DAHLER-LARSEN, Peter. From Programme Theory to Constructivism: On Tragic, Magic, and Competing Programmes. **Evaluation**, Vol. 7, n. 3 (2001), p. 331-349.

DAHLER-LARSEN, Peter. **The Evaluation Society**, Stanford, CA: Stanford University Press, 2012.

DALKIN, Sonia Michelle; GREENHALGH, Joanne; JONES, Diana; CUNNINGHAM, Bill; LHUSSIÉ, Monique. What's in a mechanism? Development of a key concept in realist evaluation. **Implementation Science** (2015), p. 10-49.

DEUTSCHER, Irving. Toward Avoiding the Goal-Trap in Evaluation Research. In ABT, Clark C. (Ed.). **The Evaluation of Social Programs**. Newbury Park, CA: Sage, 1976, p. 249–268.

DONALDSON, Stewart I. **Program Theory-Driven Evaluation Science: Strategies and Applications**. New York: Psychology Press, 2007.

DOUMA, Sytse; SCHREUDER, Hein. **Economic Approaches to Organizations**. Prentice Hall, 1998, (2nd edition).

DOWNS, Anthony. **Inside Bureaucracy**, Boston: Little, Brown, 1967.

DROR, Yehezkel. **Public Policymaking Reexamined**, Scranton, PA: Chandler, 1968.

DUNLEAVY, Patrick. **Democracy, Bureaucracy and Public Choice; Economic Explanations in Political Science**. UK: Harvester Wheatsheaf, Hemel Hempstead, 1991.

DUROSE, C. Revisiting Lipsky: Front-line Work in UK Local Governance. **Political Studies** Vol. 59, n. —4 (2011), p. 978–995.

DYNES, Patrick S.; MARVEL, Mary K. **Program Evaluation: An Annotated Bibliography**. London: Garland Publishing, 1987.

EASTON, David. **A Systems Analysis of Political Life**. New York: John Wiley and Sons, Inc., 1965.

ECKERBERG, Katarina. **Environmental Protection in Swedish Forestry: A Study of the Implementation Process**. Umeå: Department of Political Science, Research Report, 1987, p. 12.

ECKHOFF, Torstein. **Statens styrningsmuligheter, særlig i ressurs- og miljøspørsmål**. Oslo: Tanum-Norli, 1983.

ECKHOFF, Torstein. **Juss, moral og politikk**. Oslo: Universitetsforlaget, 1989.

EDWARDS III, George C.; SHARKANSKY, Ira. **The Policy Predicament**. San Francisco: W. H. Freeman, 1978.

EFN-rapport. Solvärme och värmepumpar: En utvärdering av det statliga energiforskningsstödet effektivitet. (Solar heat and heat pumps: An evaluation of the effectiveness of government economic support to research). Stockholm: Liber, 1985, p. 13.

ELLIS, K. Street-level Bureaucracy Revisited: The Changing Face of Frontline Discretion in Adult Social Care in England. **Social Policy & Administration** Vol. 45, n. 3 (2011), p. 221-244.

ELMORE, Richard. Organizational Models of Social Program Implementation. **Public Policy** Vol. 26, n. 2 (1978), p. 185-228.

ELMORE, Richard. Backward Mapping: Implementation Research and Policy Decisions. **Political Science Quarterly** Vol. 94, n. 4 (1980), p. 601-616.

ELSTER, Jon. **Logic and Society**. Chichester: Wiley, 1978.

ELSTER, Jon. **Explaining Social Behavior: More Nuts and Bolts for the Social Sciences**. Cambridge: Cambridge University Press, 2007.

ESPING, Hans, *Ramlagar i förvaltningspolitiken* (Framework laws in administrative public policy), Stockholm: SNS, 1994.

ETZIONI, Amitai. **A Comparative Analysis of Complex Organizations: On Power, Involvement, and Their Correlates**. New York: Free Press, 1975.

FAIRWEATHER, George W.; TORNATZKY, Louis G. **Experimental Methods for Social Policy Research**. Oxford: Pergamon Press, 1977.

FERNÁNDEZ-BALLESTEROS, Rocío. A Model for Planning Evaluation Research. In MAYNE, John; BEMELMANS-VIDEC, Marie-Louise, *et al.* (Eds.). **Advancing Public Policy Evaluation: Learning from International Experiences**. Amsterdam: Elsevier 1992a, p. 205-213.

FERNÁNDEZ-BALLESTEROS, Rocío. **Introducción a la Evaluación Psicológica I-II**. Madrid: Ediciones Pirámide, S.A, 1992b.

FITZ-GIBBON, Carol Taylor; MORRIS, Lynn Lyons. Theory-Based Evaluation. **Evaluation Comment** (1975), p. 1-14.

FITZ-GIBBON, Carol Taylor; MORRIS, Lynn Lyons. **How to Design a Program Evaluation**. Beverly Hills: Sage, 1978. (Segunda edição, paperback, Thousand Oaks, EUA, 1988).

FORSS, Kim. **Utvärdering som hantverk: Bortom mallar och manualer**. Lund: Studentlitteratur, 2007.

FRANKE-WIKBERG, Sigbrit; LUNDGREN, Ulf P. **Att värdera utbildning: Del 1: En introduktion till pedagogisk utvärdering**. Stockholm: Wahlström & Widstrand, 1980.

FUNK, William F.; SEAMON, Richard H. **Administrative Law: Examples and Explanations**. Aspen, CO: Law & Business, 2006 (2nd Edition).

FURUBO, Jan-Eric. **Samhällsinformation och hur den kan utvärderas**. I RIKSREVISIONSVERKET *Att mäta effekter*, Stockholm: RRV 1996:6, p. 51-91.

FURUBO, Jan-Eric. Why Evaluations Sometimes Can't Be Used—and Why They Shouldn't. In RIST, Ray C.; STAME, Nicoletta (Eds.). **From Studies to Streams: Managing Evaluative Systems**. New Brunswick, NJ: Transaction, 2006, p. 147-165.

FURUBO, Jan-Eric; RIST, Ray C.; SANDAHL, Rolf (Eds.). **International Atlas of Evaluation**. New Brunswick, NJ: Transaction Books. 2002.

GALTUNG, Johan. **Fengselssamfunnet: Et forsøk på analyse**. Oslo: Universitetsforlaget, 1959.

GEORGE, Alexander L; BENNETT, Andrew. **Case Studies and Theory Development in the Social Sciences**. Cambridge, MA: MIT Press. 2005.

GERRING, John; BARRESI, Paul A. (Eds.). Joining Minimal Definitions and Ideal Types. In COLLIER, David; GERRING, John (Eds.). **Concepts and Method in Social Science: The Tradition of Giovanni Sartori**. London: Routledge, 2009. P. 241-268.

GJELSTRUP, Gunnar. **Offentlige elbesparelser: teknik, organisation og politik** (Public electricity savings: technology, organization and politics). København: AKF Forlag, 1991.

GLENNAN, Stuart. Mechanisms and the Nature of Causation. **Erkenntnis**, Vol 44, n.1 (1996), p. 49-71

GOLDSMITH, Stephen; EGGERS, William D. **Governing by Network: The New Shape of the Public Sector**. The Brookings Institution, Washington, DC, 2004.

GREENE, Jennifer C. Book review of VEDUNG, Evert, **Public policy and program evaluation**. (Transaction 1997.) **Administrative Science Quarterly**, Vol. 44, n. 2 (1999), p. 433-436.

GUBA, Egon G; LINCOLN, Yvonna S. **Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches**, San Francisco: Jossey-Bass, 1981.

GUBA, Egon G.; LINCOLN, Yvonna S. **Fourth Generation Evaluation**. Newbury Park, CA: Sage, 1989.

HADENIUS, K. **Jämlikhet och Frihet. Politiska mål för den svenska grundskolan. PhD Thesis** (published). Uppsala: Acta Universitatis Upsaliensis, Political Science Association of Uppsala, Nr. 110, 1990.

HAMMOND, Michael; WELLINGTON, Jerry J. **Research methods: The Key Concepts**. New York: Routledge, 2013.

HANF, Kenneth. Regulatory Structures: Enforcement as Implementation. **European Journal of Political-Research** Vol. 10, n. 2 (June 1982), p. 159-172.

HANF, Kenneth; TOONEN, Theo A. J. (Eds.). **Policy Implementation in Federal and Unitary Systems**. Dordrecht: Martinus Nijhoff Publishers, 1985.

HANSEN, Morten Balle; VEDUNG, Evert. **Common Language in the Organization of Eldercare: Evaluation of a Standardized Classification System** [in Danish: Fælles sprog i ældreplejens organisering: Evaluering af et standardiseret kategorisystem]. Odense: Syddansk Universitetsforlag, 2005.

HANSEN, Morten Balle; VEDUNG, Evert. Theory-Based Stakeholder Evaluation. **American Journal of Evaluation** Vol. 31, n. 3 (2010), p. 295-313.

HARALDSSON, Jennie; HOLMER, Jan; KETOLA, Katja; KRANTZ, Lars-Gunnar. Tid för gemensam-reflektion: Självvärderingar inom storstadssatsningen i Göteborg. Göteborg: Göteborg City, Stadskansliet and Göteborg Universitet. Department of Working Life Science, 2005.

HARGROVE, Erwin C. **The Missing Link: The Study of Implementation of Social Policy**. Washington, DC: Urban Institute, 1975.

HARGROVE, Erwin C. The Search for Implementation Theory. In ZECKHAUSER Richard; LEEBAERT, D. (Eds.). **The Role of Government in the 1980's**. Durham, NC: Duke University Press, 1983, p. 280-294.

HATRY, Harry P; WINNIE, Richard E; FISK, Donald M. **Practical Program Evaluation for State and Local Governments**. Washington, DC: The Urban Institute Press, 1981 (2nd edition).

HAYEK, Friedrich A. **New Studies in Philosophy: Politics, Economics, and the History of Ideas**. Chicago: University of Chicago Press, 1978.

HAYEK, Friedrich A. **The Counter Revolution of Science**. Indianapolis, Ind: Liberty Press, 1979 (2nd edition).

HECKSCHER, Gunnar. **Svensk statsförvaltning i arbete**. Stockholm: SNS förlag, 1958.

HELLSTERN, Michael; WOLLMANN, Hellmut. (Eds.). **Experimentelle Politik: Reformstrohfeuer oder Lernstrategie, Bestandsaufnahme und Evaluierung**. Opladen: Westdeutscher Verlag, 1983.

HEMENWAY, David. **Monitoring and Compliance: the Political Economy of Inspection**, Greenwich, CN: JAI Press, 1985.

HENRY, Gary T.; JULNES, George; MELVIN, Mark M. (Eds.) **Realist Evaluation: An Emerging Theory in Support of Practice**. New Directions for Evaluation, 78. San Francisco: Jossey-Bass, 1998.

HERMERÉN, Göran; ÅSTRÖM, Karsten. Rapport R25; Stympedel i samhällsplaneringen. Stockholm. Statens råd för byggnadsforskning. Svensk Byggjänst, 1982. 63s

HERMAN, Joan L. (Ed.). **Program Evaluation Kit**. Newbury Park, CA: Sage, 1987 (2nd edition). [O kit se compõe dos seguintes volumes: 1. Evaluator's Handbook, 2. How to Focus Evaluation, 3. How to Design Program Evaluation, 4. How to Use Qualitative Methods in Evaluation, 5. How to Assess Program Implementation, 6. How to Measure Attitudes, 7. How to Measure Performance and Use Tests, 8. How to Analyze Data, 9. How to Communicate Evaluation Findings.]

HERTTING, Nils. **Samverkan på spel: Rationalitet och frustration nätverksstyrning och svensk-stadsdelsförnyelse**. [Games Local Actor Play: Rationality and frustration in network governance and Swedish neighbourhood renewal.]. Stockholm, 2003.

HERTTING, Nils; VEDUNG, Evert. **Den utvärderingstäta politiken: Styrning och utvärdering i svensk storstadspolitik**. Lund: Studentlitteratur, 2009.

HERTTING, Nils. Implementering: perspektiv och mekanismer. In ROTHSTEIN, Bo (Ed.). **Politik som organisation: förvaltningspolitikens grundproblem**. Lund: Studentlitteratur, 2015, p. 185-212.

HILDÉN, Mikael; LEPOLA, Jukka; MICKWITZ, Per; MULDER, Aard; PALOSAARI, Marika; SIMILÄ, Jukka; SJÖBLOM, Stefan; VEDUNG, Evert. **Evaluation of Environmental Policy Instruments: A Case Study of the Finnish Pulp & Paper and Chemical Industries**. Helsinki: Finnish Environment Institute, Monographs of the Boreal Environmental Research, 21, 2002.

HIRSCHMAN, Albert. **The Rhetoric of Reaction: Perversity, Futility, Jeopardy**. Cambridge, MA: Harvard University Press, 1991.

HJERN, Benny. Implementation Research – The Link Gone Missing. **Journal of Public Policy** Vol. 2, n. 3 (1982), p. 301-308.

HJERN Benny; LUNDMARK, Kjell. Blandekonomi och postweberiansk förvaltningsteori: Administrativa lärdomar av svensk regionalpolitisk forskning. **Statsvetenskapligt Tidskrift** n. 4 (1979), p. 257-266.

HJRTSJÖ, Marie. **Med samarbete i sikte: om samordnade insatser och samlokaliserade familjecentraler**. Lund: Lund University, Dept of Social Work, 2005.

HOLMER, Jan; NELSSON Olof; NILSSON, Karl-Axel; ROVIO-JOHANSSON, Airi; ÅBERG, Jan-Olof. **Självvärdering och extern granskning: utvärdering av tre utbildningar**. UHÄ-FoU projektrapport, 1992:3.

HOOD, Christopher C. **The Tools of Government**. London: Macmillan, 1983.

HOOD, Christopher C.; MARGETTS, Helen Z. **The Tools of Government in the Digital Age**, Houndmills, Hampshire: Palgrave Macmillan, 2007.

HOOPERWERF, Andries. Reconstructing Policy Theory. **Evaluation and Program Planning**, Vol. 13, n. 3 (1990), p. 285-291.

HOOPERWERF, Andries. Policy Evaluation and Government in the Netherlands: Meta Evaluation Research as One of the Solutions. In MAYNE, John; BEMELMANS-VIDEC, Marie-Louise *et al.* (Eds.). **Advancing Public Policy Evaluation: Learning from International Experiences**. Amsterdam: Elsevier, 1992, p. 215-227.

HOUSE, Ernest R. **Evaluating with Validity**. Newbury Park, CA: Sage, 1980.

HOUSE, Ernest R. **Professional Evaluation: Social Impact and Political Consequences**. Newbury Park, CA: Sage, 1993.

HOUSE, Ernest R. Democracy and Evaluation. Paper presented to the *European Evaluation Society*, Berlin, 2 October 2004.

HØYBYE-MORTENSEN, M. Velfærdstatens dørvogtere: Procesregulering af visitationer på socialområdet [Gatekeepers of the welfare state: Administrative procedures in social work decision-making]. 2011. Copenhagen: Department of Political Science, University of Copenhagen.

HUDSON, Joe; MAYNE, John; THOMLISON, Ray (Eds.). **Action-Oriented Evaluation in Organizations: Canadian Practices**. Toronto, Ontario: Wall & Emerson, 1992.

IMAI, M. **Kaizen: The Key To Japan's Competitive Success**. New York: McGraw-Hill. 1986.

IMAS, Linda G. Morra; RIST, Ray C. **The Road to Results: Designing and Conducting Effective Development Evaluations**, Washington DC: World Bank, 2009.

JEWELL, C. **Agents of the Welfare State. How Caseworkers Respond to the Need in the United State, Germany, and Sweden**. US: Palgrave, Macmillan US, 2007.

JOHANSSON, Jan. **Det statliga kommittéväsendet: Kunskap, kontroll, konsensus**, Stockholm: Stockholm University, Department of Political Science, Stockholm Studies in Politics 41, 1992.

JONUNG, Lars (with Jan RYDENFELT) **Prisregleringen, företagen och förhandlingsekonomin**, Stockholm: SNS, 1984.

JØRGENSEN, Torben Beck. **Når staten skal spare**. Copenhagen: Nyt fra samfundsvidenskaberne, 1981.

JUDD, Charles M.; KENNY, David A. **Estimating the Effects of Social Interventions**. Cambridge: Cambridge University Press, 1981.

KARLSSON, Christer. **Democracy, Legitimacy and the European Union**. Uppsala: Uppsala University, Department of Government, Acta Universitatis Upsaliensis, 146. PhD thesis, 2001.

KARLSSON, Ove. A Critical Dialogue in Evaluation: How can the Interaction Between Politics and Evaluation Be Tackled? **Evaluation** Vol. 2, n. 4 (1996), p. 405-416.

KARLSSON, Ove. Socratic Dialogue in the Swedish Political Context. **New Directions for Evaluation** Vol. 77 (1998), p. 21-38.

KARLSSON, Ove. Critical Dialogue: Its Value and Meaning. **Evaluation** Vol. 7, n. 2 (2001), p. 211-227.

KAUFMAN, Herbert. **The Forest Ranger: A Study in Administrative Behavior**. Baltimore: Johns Hopkins Press, 1967.

KAUFMAN, Roger; THOMAS, Susan. **Evaluation without Fear**. New York: New Viewpoints 1980.

KELMAN, Steven. **Regulating America, Regulating Sweden: A Comparative Study of Occupational Safety and Health Policy**. Cambridge, MA: MIT Press, 1981.

KIVIMAA, Paula; MICKWITZ, Per. The Challenge of Greening Technologies – Environmental Policy Integration in Finnish Technology Policies. **Research Policy** Vol. 35, n. 5 (2006), p. 729-744.

KIVIMAA, Paula. **The Innovation Effects of Environmental Policies: Linking Policies, Companies, and Innovations in the Nordic Pulp and Paper Industry**. Helsinki: PhD Thesis, Helsinki School of Economics, A-329, 2008.

KLINGBERG, Tage. **Val och användning av styrmedel: En studie av byggnadsnämndens tillsyn**, (Selection and use of public policy instruments: A study of the Building Commissions' inspections). Stockholm: Svensk Byggtjänst, BFR skrift T10: 1980.

KOSECOFF, Jacqueline; FINK, Arlene. **Evaluation Basics: A Practitioner's Manual**. Newbury Park, CA: Sage, 1982.

KROGSTRUP, Hanne Kathrine; JENSEN, Julie Borup. Fra New Public Management til New Public Governance. In KROGSTRUP, Hanne Kathrine (ed.). **Samskabelse og capacity building i den offentlige sektor**. Hans Reitzels Forlag, 2017, p. 33-55.

KUSEK, Jody Zall; RIST, Ray C. **Ten Steps to a Results Based Monitoring and Evaluation System**. Washington, DC: International Bank for Reconstruction and Development / World Bank, 2004.

LAFFERTY, W.; hoyden, E. Environmental policy Integration. Toward an Analytical Framework. **Environmental Politics** Vol. 12, n.3 (2003), p. 1 -22.

LANE, Jan-Erik. Implementation, Accountability, and Trust. **European Journal of Political Research** Vol. 15, n. 5 (1987), p. 527–546.

LANGE, M. **Comparative Historical Methods**. London: Sage, 2013.

LAUNSON, Laila; RIEPER, Olaf. **Forskning om og med mennesker: Forskningstyper og forskningsmetoder i samfundsforskning**. København: Nyt Nordisk Forlag Arnold Busck, 2005 (5th edition).

LEWIN, Kurt. **The Dynamic Theory of Personality**. New York: McGraw Hill, 1935.

LEWIN, Kurt. **Principles of Topological Philosophy**. New York: McGraw Hill, 1936.

LEWIN, Leif. **Det gemensamma bästa: Om egenintresset och allmänintresset i västerländsk politik**, Stockholm: Carlssons, 1988.

LIGHT, Richard J.; PILLEMER, David B. **Summing Up: The Science of Reviewing Research**, Cambridge, MA: Harvard University Press, 1984.

LINCOLN, Yvonna S.; GUBA, Egon. G. The Roots of the Fourth Generation Evaluation: Theoretical and Methodological Origins. In ALKIN, Marvin C. (Ed.). **Evaluation Roots: tracing theorist's views and influences**. Thousand Oaks, CA: Sage, 2004, p. 225-241.

LIPSKY, Michael. **Street-level Bureaucracy. Dilemmas of the individual in public services**. New York: Russell Sage Foundation, 1980. (Reeditado em 2010).

LOVE, Arnold. **Internal Evaluation: Building Organizations from Within**. Newbury Park, CA: Sage, 1991.

LUNDBERG, L. **Från lag till arbetsmiljö**. (From law to work environment) Malmö. Liber, 1982

LUNDQUIST, Lennart. Några synpunkter på begreppet politisk planering. **Statsvetenskaplig Tidskrift** (1976), p.121-139.

LUNDQUIST, Lennart. From Order to Chaos: Recent Trends in the Study of Public Administration. In LANE, J-E. (Ed.). **State and Market: The Politics of the Public and the Private**. Newbury Park, CA.: Sage 1985.

LUNDQUIST, Lennart. **Implementation Steering: An Actor-Structure Approach**, Lund: Student-litteratur, 1987.

LUNDQUIST, Lennart. **Kriterier för utvärdering av offentlig verksamhet**, (Criteria for evaluating public activities) Turku: Meddelanden från Ekon. Statsvet. Fakulteten vid Åbo Akademi, ser. A: 305. 1990.

LUNDQVIST, Lennart J.. **The Hare and the Tortoise: Clean Air Policies in the US and Sweden**, Ann Arbor, MI: University of Michigan Press, 1980.

LUNDQVIST, Lennart J; ELANDER, Ingemar; DANEMARK, Berth. Housing Policy in Sweden – still a success story? **International Journal of Urban and Regional Research**, Vol. 14, n. 3 (1990), p. 445-467.

MADAUS, George F; SCRIVEN, Michael; STUFFLEBEAM, Daniel (Eds.). **Evaluation Models: Viewpoints on Educational and Human Services Evaluation**. The Hague: Kluwer-Nijhoff, 1983.

MAJONE, Giandomenico. **Evidence, Argument and Persuasion in the Policy Process**. London: Yale University Press, 1989.

MANN, D H. Notes and Comments: A Matrix Technique for Finite Bayesian Decision Problems. **Decision Sciences** Vol. 3, n. 4 (October 1972), p. 129-136.

MARCH, James G.; OLSEN, Johan P. (Eds.). **Ambiguity and Choice in Organizations**. Bergen: Universitetsforlaget, 1976.

MARK, Melvin M.; HENRY, Gary T; JULNES, George. **Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs**. San Francisco: Jossey-Bass, 2000.

MARRA, Mita. Some Insights From Complexity Sciences for the Evaluation of Complex Policies. In FORSS, Kim; MARRA, Mita; SCHWARTZ, Robert. **Evaluating the Complex: Attribution, Contribution and Beyond**. New Brunswick, NJ: Transaction, 2011.

MATHIESEN, Thomas. **Silently Silenced – Essays on the Creation of Acquiescence in Modern Society**. Winchester: Waterside Press, 2004.

MATHISON, Sandra. **Encyclopedia of Evaluation**. SAGE Publications: 2005.

MAYNTZ, Renate. Executive Leadership in Germany: Dispersion of Power or 'Kanzler-demokratie'. In ROSE, Richard; SULEIMAN, Ezra N. (Eds.). **Presidents and Prime Ministers**. Washington, DC: American Enterprise Institute, 1980.

MAYNTZ, Renate (Ed.) **Implementation Politischer Programme: Ansätze zur Theoriebildung**. Opladen: Westdeutscher Verlag, 1982.

MAZMANIAN, Daniel A; SABATIER, Paul A. **Effective Policy Implementation**. Lexington, MA: Lexington Books, 1981.

MERTON, Robert K. **Social Theory and Social Structure**. London: Collier-Macmillan, 1968.

MEIER, K.J. **Regulation: Politics, Bureaucracy, and Economics**. New York: St. Martin's, 1985.

MEYERS, William R. **The Evaluation Enterprise: A Realistic Appraisal of Evaluation Careers, Methods, and Applications**. San Francisco: Jossey-Bass, 1981.

MILLER, George. The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. **Psychological Review** Vol. 101, n. 2 (1956), p. 343-252.

MICKWITZ, Per. **Environmental Policy Evaluation: Concepts and Practices**. Helsinki: Finnish Society of Sciences and Letters, 2006.

MICKWITZ, Per; KIVIMAA, Paula. Evaluating Policy Integration: The Case of Policies for Environmentally Friendlier Technological Innovations. **Evaluation** Vol. 13, n. 1 (2007), p. 68-86.

MOHR, Lawrence B. **Impact Analysis for Program Evaluation**. Newbury Park, CA: Sage, 1995.

MONTJOY, Robert S; O'TOOLE, Laurence J. Towards a Theory of Policy Implementation: An Organizational Perspective. **Public Administration Review** Vol. 39, n. 5 (1979), p. 465-476.

MUNDACA TORO, Luis Alberto. **Markets for Energy Efficiency: Exploring the New Horizons of Tradable Certificate Schemes**. Lund: Lund University, International Institute for Industrial Environmental Economics, 2008.

MUNICIO, Ingegerd. Implementationsforskning: En litteraturöversikt. **Statsvetenskaplig Tidskrift** Vol. 85, n. 3 (1982), p. 183-190.

MUNICIO, Ingegerd. **Fran lag till bruk : hemspraksreformens genomförande**. Stockholm: Centrum for invadringforskning, Stockholm studies in politics 31.Thesis (Doctoral) Stockholms universitet, 1987.

MURRAY, Charles. **Losing Ground: American Social Policy, 1950–1980**. New York: Basic Books, 1984.

NACHMIAS, David. **Public Policy Evaluation: Approaches and Methods**. New York: St. Martin's, 1979.

NAKAMURA, Robert T.; SMALLWOOD, Frank. **The Politics of Policy Implementation**, New York: St. Martin's, 1980.

NIELSEN, Birgitta Gomez. **The Hidden Politics of a Haunted Sector: Retrenchment in Danish Housing Policy 2001-2009**. Copenhagen: University of Copenhagen, Dept of Political Science, 2010.

NIELSEN, Steffen Bohni; JACOBSEN, Mads Nyholm; PEDERSEN, Morten. Øje for effekterne – resultatbaseret styring kan styrke offentlige indsatser. **Nordisk Administrativ Tidsskrift**, årg. 86, nr. 4, 2005, p. 276-295.

NILSSON, M; ECKERBERG, K. **Environmental Policy Integration in Practice**. Shaping Institutions for Learning. London: Earthscan, 2007.

NILSTUN, Tore; HERMERÉN, Göran. **Utvärderingsforskning och rättsliga reformer: Analys av orsaker och effekter**. Lund: Studentlitteratur, 1984.

NISKANEN, William A. **Bureaucracy and Representative Government**, Chicago: Aldine-Atherton, 1971.

NUTLEY, S. M.; WALTER, I.; DAVIES, H. T. O. **Using Evidence: How Research Can Inform Public Services**. Bristol: Policy Press, 2007.

OECD/DAC. **Glossary of Key Terms in Evaluation and Results-Based Management**, Paris: OECD Publications, Development Assistance Committee (DAC). 2002. <https://www.oecd.org/dac/evaluation/2754804.pdf>

OLLAIK, Leila Giandoni; MEDEIROS, Janann Joslin. Instrumentos governamentais: Reflexões para uma agenda de pesquisas sobre implementação de políticas públicas no Brasil. **Revista de Administração Pública** Vol 45, n. 6 (Dez 2011), p. 1943-1967.

ORWELL, George. Politics and the English Language. In **The Collected Essays, Journalism, and Letters of George Orwell, Volume IV, In Front of your Nose 1945–1950**. Harmondsworth, Middlesex: Penguin Books, 1970, P. 156–170.

OSBORNE, Stephen P. The New Public Governance? **Public Management Review** Vol. 8, n 3, (2006), pp 377-387.

OWEN, John M; ROGERS, Patricia J. **Program Evaluation: Forms and Approaches**. St. Leonards, Australia: Allen & Urwin, 1998.

PATTON, Michael Q. **How to Use Qualitative Methods in Evaluation**, Newbury Park, CA: Sage, 1987.

PATTON, Michael Q. **Creative Evaluation**. London: Sage, 1987.

PATTON, Michael Q. **Qualitative Evaluation and Research Methods**, Newbury Park, CA: Sage, 1990 (2nd edition).

PATTON, Michael Q. **Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use**. London: Guilfords, 2011.

PAWSON, Ray. **The Science of Evaluation. A Realist Manifesto**. London: Sage, 2013.

PEDERSEN, C. S.; DAHLER-LARSEN, Peter; ANDERSEN, V. N.; HANSEN, K. M. (Eds.). **Selvevalueringens hvide sejl**. Odense: Syddansk Universitetsforlag, 2003.

PESTOFF, Victor; BRANDSEN, T.; VERSCHUERE, B. (Eds.). **New Public Governance, the Third Sector and Co-Production**. Routledge, 2012.

PETERSSON, Gustav Jakob. **Program Theory Evaluation of Public-private Agreements on Environmental Programmes and Projects**. In EASY-ECO CONFERENCE, 11-14 October 2006, Saarbrücken, Germany.

PIHLGREN, Gunnar; SVENSSON, Arne. **Målstyrning: 90-talets ledningsform för offentlig verksamhet**. Malmö: Liber/Hermods, 1989.

PIERSON, Paul. **Dismantling the Welfare State? Reagan, Thatcher and the Politics of Retrenchment**. Cambridge: Cambridge University Press. 1994.

POLLITT, Christopher. **The Essential Public Manager**, Maidenhead, UK: Open University Press, 2003.

POPPER, Karl R. **The Poverty of Historicism**. New York: Harper & Row, 1961. Tradução em português: **A miséria do historicismo**. São Paulo: EDUSP, 1980.

POPPER, Karl R. The Rationality Principle. In MILLER, David, (Ed.). **Popper Selections**, Princeton, NJ: Princeton University Press, 1985 (first published in 1967), p. 357-365.

POPPER, Karl R. **The Open Society and Its Enemies**. London: Routledge, 1962. Tradução em português: **A sociedade aberta e seus inimigos**. São. Paulo: Ed. da Universidade de São Paulo, 1974. 2 Vols. (Belo Horizonte: Itatiaia, 1998, 2ª edição).

POPPER, Karl R. **A Lógica da Pesquisa Científica**. São. Paulo: Ed. Cultrix, 1972. (2ª edição em 2013).

POPPER, Karl R. **Conjectures and Refutations: The Growth of Scientific Knowledge**. London: Routledge & Kegan Paul, 1974 (5th edition). Tradução em português: **Conjecturas e Refutações**. Brasília, Editora da UnB, 1980).

POWER, Michael. **The Audit Society: Rituals of Verification**. Oxford: Oxford University Press, 1997.

PREMFORS, Rune. **Policyanalys: Kunskap, praktik och etik i offentlig verksamhet**. Lund: Studentlitteratur, 1989.

PRESSMAN, Jeffrey; WILDAVSKY, Aaron B. **Implementation: How Great Expectations in Washington are Dashed in Oakland etc.** Berkeley, CA: University of California Press, 1984 (3rd edition).

PRIOR, Lindsay. **Using Documents in Social Research**. SAGE Publications Ltd. Series: Introducing Qualitative Methods. 2003.

RIPLEY, Randall B.; FRANKLIN, Grace A. **Policy Implementation and Bureaucracy**. Chicago, IL: Dorsey, 1986 (2nd edition).

RIST, Ray C. (Ed.). **Program Evaluation and the Management of Government: Patterns and Prospects across Eight Nations**. London: Transaction Publishers, 1990.

RIVLIN, Alice. **Systematic Thinking for Social Action**, Washington DC: Brookings, 1971.

ROMBACH, H. D. Practical Benefits of Goal-oriented Measurement. In FENTON, N.; LITTLEWOOD, B. (Eds.). **Software Reliability and Metrics**, London: Elsevier Applied Science, 1991.

ROSENTHAL, Robert. **Meta-Analytical Procedures for Social Research**. Newbury Park, CA: Sage, 1984.

ROSSI, Peter H; FREEMAN, Howard; LIPSEY, Mark W **Evaluation: A Systematic Approach**. London: Sage, 1985, 1989, 1993 (3rd, 4th, 5th editions). With Mark Lipsey as co-author starting in 1999 (6th edition)

ROTHSTEIN, Bo. **Väd bör staten göra. Den generella välfärdspolitikens moraliska och politiska logik** (What should the State do? The moral and political logic of the general welfare policy). Stockholm: SNS Forlag, 1994.

ROTHSTEIN, Bo. Sociala fällors och tillitens problem. In HOLMBERG, Sören; WEIBULL, Lennart. (Eds.). **Ett missnöjt folk?** Göteborg: SOM-institutet, Göteborg Universitet. 1997.

RUSSELL, Bertrand. **Mysticism and Logic**. 1918. Sobre a Noção de Causa. In **Misticismo e Lógica**. Tradução de Wilson Velloso. São Paulo: Companhia Editora Nacional, 1957.

RUTMAN, Leonard (Ed.) **Evaluation Research Methods: A Basic Guide**. Beverly Hills, CA: Sage, 1977.

RUTMAN, Leonard. **Planning Useful Evaluations: Evaluability Assessment**. Newbury Park, CA: Sage, 1980.

RUTMAN, Leonard; MOWBRAY, George. **Understanding Program Evaluation**, Newbury Park, CA: Sage, 1983.

RYDÉN, Bengt (Ed.). **Makt och vanmakt: Lärdomar av sex borgerliga regeringsår**. Stockholm: SNS, 1983.

RRV - **Riksrevisionsverket**, Stockholm, 1985, 1986.

SABATIER, Paul; MAZMANIAN, Daniel. The Implementation of Public Policy: A Framework of Analysis. **Policy Studies Journal** Vol. 8, n. 4 (1980), p. 538-560.

SABATIER, Paul; MAZMANIAN, Daniel. The Implementation of Public Policy: A Framework for Analysis. In MAZMANIAN, Daniel; SABATIER, Paul. (Eds.). **Effective Policy Implementation**. Lexington, MA: D.C. Heath, 1981.

SAETREN, Harald. **Iverksetting av offentlig politikk: en studie av utflytting av statsinstitusjoner fra Oslo 1960-1981**. Universitetsforlaget, 1983.

SAETREN, Harald. Facts and Myths about Research on Public Policy Implementation: Out-of-Fashion, Allegedly Dead, But Still Very Much Alive and Relevant. **Policy Studies Journal** Vol. 33, n. 4 (2005), p. 559-582.

SAETREN, Harald. Implementing the Third Generation Research Paradigm in Policy Implementation Research: An Empirical Assessment. **Public Policy and Administration** Vol. 29, n. 2 (2014), p. 84-105.

SAHLIN-ANDERSSON, Kerstin. **Oklarhetens strategi. (The strategy of obscurities)** Lund: Studentlitteratur, 1989.

SALAMON, Lester M. Rethinking Public Policy: Third Party Government and the Changing Forms of Government Action. **Public Policy** Vol. 29, n. 3 (1981), p. 255-275.

SALAMON, Lester M.; LUND, Michael S. The Tools Approach: Basic Analytics. In SALAMON, Lester M. (Ed.). **Beyond Privatization: The Tools of Government Action**. Washington, DC: Urban Institute Press, 1989, p. 23-49.

SANDBERG, Bo; FAUGERT, Sven. **Perspektiv på utvärdering**. Lund: Studentlitteratur, 2007.

SANDAHL, Rolf. **Instruments of Swedish Housing Policy**. PhD thesis, Uppsala University. 1983.

SANDAHL, Rolf; PETERSSON, Gustav Jakob. **Kausalitet: I filosofi, politik och utvärdering**. Lund: Studentlitteratur, 2016.

SARTORI, Giovanni. **Parties and Party Systems: A Framework for Analysis, Vol. I**. Cambridge: Cambridge University Press, 1976.

SARTORI, Giovanni. Guidelines for Concept Analysis. In SARTORI, Giovanni. (Ed.). **Social Science Concepts: A Systematic Analysis**. Beverly Hills: Sage, 1984, p.15-85. Also in COLLIER, David; GERRING, John (Eds.). **Concepts and Method in Social Science: The Tradition of Giovanni Sartori**. London: Routledge, 2009a, p. 97-150.

SARTORI, Giovanni. Tower of Babel. In COLLIER, David; GERRING, John (Eds.). **Concepts and Method in Social Science: The Tradition of Giovanni Sartori**. London: Routledge, 2009b, p. 61-96.

SCHIERENBECK, Isabell. Bakom välfärdsstatens dörrar (Behind the doors of the welfare state). PhD dissertation, Göteborg: University of Gothenburg, Political Science Dept., 2003.

SCHLAEGER, Jesper, **The World in Evaluation? Historical Contingency: Irony and Comedy** (in Danish), Copenhagen: Copenhagen University, Institute of Political Science, assignment in the seminar Evaluation in the Public Sector (Vedung). 2006

SCHLAEGER, Jesper. **E-Government in China: Technology, Power and Local Government Reform**, London: Routledge, 2013.

SCRIVEN, Michael. Goal-Free Evaluation. In HOUSE, Ernest R. (Ed.). **School Evaluation: The Politics and Process**. Berkeley, CA: McCutchan, 1973.

SCRIVEN, Michael. Pros and Cons About Goal-Free Evaluation. **Evaluation Comment** Vol. 3 (1974), p. 1–4.

SCRIVEN, Michael. **The Logic of Evaluation**. Inverness, CA: Edgepress, 1980.

SCRIVEN, Michael. **Evaluation Thesaurus**. Newbury Park, CA: Sage, 1991.

SCRIVEN, Michael. Evaluation in the New Millenium: The Transdisciplinary Vision. In DONALDSON, Stewart I.; SCRIVEN, Michael. (Eds.). **Evaluating Social Programs and Problems: Visions for the New Millenium**. Mahwah, NJ; Lawrence Erlbaum Associates, 2003, p. 19-42.

SEVÄ, Mikael. The Decisive Role of Street-level Bureaucrats in Environmental Management. Doctoral dissertation. Luleå: Luleå Tekniska Universitet, PhD-dissertation, 2015.

SHADISH Jr., William R.; COOK, Thomas D.; LEVITON, Laura C. **Foundations of Program Evaluation: Theory and Practice**. London: Sage, 1991.

SHAW, George Bernard. In *Oxford Dictionary of Quotations*. PARTINGTON, Angela (Ed.) Oxford: Oxford University Press, 1992, 636:26 (quote from 1907).

SIDA- SWEDISH INTERNATIONAL DEVELOPMENT COOPERATION AGENCY. Looking Back, Moving Forward. Stockholm: Sida Evaluation Manual, 2007 (2nd revised edition).

SIEBER, Sam D. **Fatal Remedies: The Ironies of Social Intervention**. New York: Plenum Press, 1981.

SIMILÄ, Jukka. **Regulating Industrial Pollution: The Case of Finland**. FORUM IURIS Helsingin yliopiston oikeustieteellisen tiedekunnan julkaisut. Helsinki, 2007.

SIMON, Herbert A. **Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations**. London: Collier-Macmillan, 1976 (3rd edition). Publicado em português: SIMON, Herbert. **Comportamento Administrativo**. Rio de Janeiro: FGV, 1979. SJÖBLOM, Stefan. **Kostnadsvariationer i kommunal serviceproduktion: Det allmänna biblioteksväsendet**. Turku: Åbo Akademi University: Meddelanden från Ekon.-statsvet. Fakulteten, Ser. A: 355, 1991.

SKOGSTYRELSENS FÖRFATTNINGSSAMLING, (The Forestry Act). SKSFS. Skogsstyrelsen (The National Forestry Board). Jönköping. 1986.

SMITH, Adam. **Wealth of Nations.** New York: Random House. 1937 (1776).

SONNICHSEN, Richard. C. **High Impact Internal Evaluation.** Thousand Oaks, CA: Sage, 2000.

SÖDERLIND, D.; PETERSSON, O. **Svensk förvaltningspolitik.** Uppsala: Diskurs, 1988.

SOU 1983:34, STATENS OFFENTLIGA UTREDNINGAR (Official publications of Sweden's government office). **Information som styrmedel: En rapport om energisparkommitténs verksamhet och energihushållningsprogrammets effekter,** Stockholm: Liber.

STAKE, Robert E. **Evaluating the Arts in Education: A Responsive Approach.** Columbus, OH: Merrill, 1975.

STATENS ENERGIVERK. **Energihushållningsprogrammets effekter: en analys av de stat-liga styrmedlens effekter på energihushållningen,** 1984:2, Stockholm: Liber.

STENQVIST, Christian. **Industrial Energy Efficiency Improvement: The role o policy and evaluation.** Lund: Lund University, Environmental and Energy Systems Studies. PhD Thesis, 2013.

STERNER, Thomas; CORIA, Jessica. **Policy Instruments for Environmental and Natural Resource Management.** New York: Resources for the Future Press, 2011 (2nd edition).

STJERNQUIST, Per. **Laws in the Forest: A Study of Public Direction of Swedish Private Forestry.** Lund: CWK Gleerup, 1973.

STOCKMANN, Reinhard; MEYER, Wolfgang. **Functions, Methods, and Concepts in Evaluation.** Houndmills, Hampshire: Palgrave Macmillan, 2013.

STONE, Clarence N. The Implementation of Social Programs. **Journal of Social Issues** Vol. 36, n. 4 (1980), p. 13–34.

STONE, Clarence N. Efficiency versus Social Learning: A Reconsideration of the Implementation Process. **Policy Studies Review** Vol. 4, n.3 (1985), p. 484–496.

STONE, Deborah. **The Policy Paradox: the Art of Political Decision Making,** London: W.W. Norton & Co. 2001 (2nd revised edition), 2012 (3rd edition).

STUFFLEBEAM, Daniel L. Evaluation Models. **New Directions for Evaluation** Vol. 89, Special issue, Spring 2001.

STUFFLEBEAM, Daniel L.; MADAUS, G. F.; KELLAGAHAN. T. **Evaluation Models**. Boston: Kluwer, 2000.

STUFFLEBEAM, Daniel L. The 21st-Century CIPP Model: Origins, Development and Use. In ALKIN, Marvin C. (Ed.) **Evaluation Roots: Tracing Theorist's Views and Influences**. SAGE, 2004, p. 245-266 (revised 2013).

STUFFLEBEAM, Daniel L.; SHINKFIELD, Anthony J. **Evaluation Theory, Models and Applications**. San Francisco, CA: Jossey-Bass 2007.

SUCHMAN, Edward A. **Evaluative Research: Principles and Practice in Public Service and Social Action Programs**. New York: Russell Sage, 1967.

SUCHMAN, Edward A. Action for What? A Critique of Evaluative Research. In WEISS, Carol H. (Ed.). **Evaluating Action Programs**. Boston: Allyn & Bacon, 1972, p. 52–84.

SUNDSTROM, Göran. **Stat pa villovagar: Resultstyrningens framvax i ett historisk-institutionellt perspektiv**. Stockholm: Stockholm University, Department of Political Science. 2003.

SUNDSTRÖM, Göran. Management by Results: Its Origin and Development in the Case of the Swedish State. **International Public Management Journal** Vol. 9, n. 4 (2006), p. 399–427.

SVENSK FÖRFATTNINGSSAMLING (SWEDISH CODES OF STATUTES). Stockholm: Liber. 1979.

TURNBULL, Barbara. Program Theory Building: A Strategy for Deriving Cumulative Evaluation Knowledge. **American Journal of Evaluation** Vol. 23, n. 3 (2003), p. 275-290.

UNDERDAL, A. Integrated Marine Policy: What? Why? How? **Marine Policy** (July 1980), p. 159-169.

VAN DER MEER, Frans-Bauke; EDELENBOS, Julian. Evaluation in Multi-Actor Policy Processes: Accountability, Learning and Co-operation. **Evaluation** Vol. 12, n. 2 (2006), p. 201-218.

VAN HORN, Carl E.; VAN METER, Donald S. The Implementation of Intergovernmental Policy. In NAGEL, Stuart S. (Ed.). **Policy Studies Review Annual**, 1. Newbury Park, CA: Sage, 1977, p. 97–120.

VAN METER, Donald S.; VAN HORN, Carl E. The Policy Implementation Process: A Conceptual Framework. **Administration and Society** Vol. 6, n. 4 (1975), p. 445–488.

VEDUNG, Evert. **Det rationella politiska samtalet: Hur politiska budskap tolkas, ordnas och prövas** (The Rational Political Conversation: Interpreting, Ordering and Assessing Political Messages). Stockholm: Bonniers, 1977.

VEDUNG, Evert. **Energipolitiska utvärderingar 1973-1981** (Energy Policy Evaluations 1973-1981). Stockholm: Liber, Delegationen för energiforskning, rapport nr. 52. 1982a.

VEDUNG, Evert. **Political Reasoning**, Newbury Park: Sage, 1982b.

VEDUNG, Evert. Systematic Interpretation and Charity. **Informal Logic Newsletter**. University of Windsor (Canada) Vol 5, no. 2 (June 1983), p. 21.

VEDUNG, Evert. **Statens markpolitik, kommunerna och historiens ironi** (Governmental Land Use Policies, the Municipalities, and the Irony of History). Stockholm: SNS Förlag, 1993, p. 131-147. <http://www.diva-portal.org/smash/get/diva2:1195887/FULLTEXT01.pdf> (accessed 20180729)

VEDUNG, Evert. **Public Policy and Program Evaluation**. New Brunswick, NJ and London: Transaction, 1997.

VEDUNG, Evert. **Utvärdering i politik och förvaltning** (Evaluation in Public Policy and Public Administration). Lund: Studentlitteratur, 1998a.

VEDUNG, Evert. Policy Instruments: Typologies and Theories. In BEMELMANS-VIDEC, Marie-Louise; RIST, Ray C.; VEDUNG, Evert. (Eds.). **Carrots, Sticks, and Sermons: Policy— Instruments and Their Evaluation**. New Brunswick, New Jersey and London: Transaction, 1998b, p. 21–58.

VEDUNG, Evert; KLEFBOM, Erik. **Ozonhålet: Globalt, nationellt, lokalt**. (The ozone hole: Global, national and local aspects). Malmö: Liber ekonomi, 2002.

VEDUNG, Evert. Styrmedel. In BOVERKET, ENERGI MYNDIGHETE & NATURVÅRDS-VERKET. (Eds), **Effektivare energi i bostäder: En antologi om framtidens styrmedel.** (Policy instruments. In: More effective energy in residential buildings: An anthology on future policy instruments) Eskilstuna: Energimyndighetens Förlag, 2002, p. 94-113.

VEDUNG, Evert. Evaluation Research and Fundamental Research. In STOCKMANN, R. (Ed.). **Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder.** Münster: Waxmann 2006, p. 113-136; Opladen: Leske + Budrich, 2000, p. 103-126.

VEDUNG, Evert; SVÄRD, Sven-Erik. Anteciperung i implementeringsprocesser. In GUSTAVSSON, Sverker; HERMANSSON, Jörgen; HOLMSTRÖM, Barry (Eds.). **Statsvetare ifrågasätter: Uppsalamiljön vid tiden för professorsskiftet den 31 mars 2008.** (Anticipation in Implementation Processes. In: Political scientists call in question: the Uppsala community at the time of the changing of professors in March 31 2008.) Uppsala: Uppsala University Library, 2008, p. 234-249.

VEDUNG, Evert. **Utvärdering i politik och förvaltning (Evaluation in Public Policy and Public Administration).** Lund: Studentlitteratur, 2009 (3rd edition).

VEDUNG, Evert. Four Waves of Evaluation. **Evaluation**, Vol. 16, n.3 (2010), p. 263-277.

VEDUNG, Evert. Side Effects, Perverse Effects and Other Strange Effects of Public Interventions. In SVENSSON, Lennart; BRULIN, Göran; JANSSON, Sven & SJÖBERG, Karin. (Eds) **Capturing Effects of Projects and Programmes**, Lund: Studentlitteratur, 2013, p 35-62.

VEDUNG, Evert; DAHLBERG, Magnus. **Demokrati och brukarutvärdering (Democracy and client-oriented evaluation)** Lund: Studentlitteratur, 2013 (2nd edition).

VEDUNG, Evert; BJURULFF, Staffan. Three Approaches to Impact Evaluation of Regional Development. In ELG, Mattias; ELLSTRÖM, Per-Erik; KLOFSTEN, Magnus & TILLMAR, Malin, (Eds), **Sustainable Development in Organizations – Studies on Innovation Practices**, Cheltenham, UK & Northampton, MA, USA: Edward Elgar, 2015, p. 313-345.—<https://www.elgaronline.com/view/9781784716882.00023.xml> (accessed 20151202)

VEDUNG, Evert. **Implementering i politik och förvaltning** (Implementation in public policy and public administration). Lund: Studentlitteratur, 2016.

VEDUNG, Evert. Indflettet og løbende udviklingsevaluering. In GHIN, Eva Moll; GRØN, Caroline Howard; KRISTIANSEN, Mads Bøge (Eds). **Styring og evaluering i den offentlige sektor: Festschrift til Hanne Foss Hansen**. (Ongoing and braided developmental evaluation. Governance and evaluation in the public sector: Essays in honor of HFH.) København: Reitzels Forlag, 2017, p 153-181.

WAGNER, Jon. (Ed.). **Images of Information**, Newbury Park, CA: Sage, 1979.

WARWICK, K.; NOLAN, A. Evaluation of Industrial Policy: Methodological Issues and Policy Lessons. In **OECD, Science, Technology and Industry Policy Papers**, No. 16, Éditions OCDE, Paris. DOI: <http://dx.doi.org/10.1787/5jz181jh0j5k-en>. 2014.

WATT BOOLSEN, Merete. **Kvalitative analyser. At finde årsager og sammenhænge**. Kopenhagen: Gyldendal Akademisk Forlag, 2006.

WEBB, Eugene J; CAMPBELL, Donald T; SCHWARTZ, Richard D; SECHREST, Lee R. **Unobtrusive Measures: Nonreactive Research in the Social Sciences**. Chicago: Rand McNally, 1966.

WEBER, Max. **From Max Weber: Essays in Sociology**. Edited with an Introduction by GERTH, Hans; MILLS, C. Wright. London: Routledge and Kegan Paul, 1974. Primeira publicação em 1948. Republicado nos anos de 1952, 57, 61, 64, 67, 70, 74.

WEISS, Carol H. **Evaluation Research: Methods of Assessing Program Effectiveness**. Englewood Cliffs, NJ: Prentice-Hall, 1972a.

WEISS, Carol H. **Evaluating Action Programs: Readings in Social Action and Education**. Boston, MA: D.C. Heath, 1972b.

WEISS, Carol H. **Organizations for Policy Analysis: Helping Government Think**. Newbury Park, CA: Sage, 1991.

WEISS, Carol. H. Theory-based Evaluation: Past, Present, and Future. **New Directions for Evaluation** issue 76 (1997), p 41-55.

WEST, William F. **Administrative Rulemaking: Politics and Processes**. Westport, CN: Greenwood Press, 1985.

WHOLEY, Joseph S.; SCANLON, John W.; DUFFY, Hugh G.; FUKUMOTO, James S.; VOGT, Leona M. **Federal Evaluation Policy: Analyzing the Effects of Public Programs**. Washington, DC: The Urban Institute, 1970.

WILDAVSKY, Aaron. If Planning Is Everything, Maybe It's Nothing. **Policy Sciences**, Vol. 4 (1973), p. 27–153.

WILDAVSKY, Aaron. **Speaking Truth to Power: The Art and Craft of Policy Analysis**. Boston: Little, Brown & Co, 1979.

WILDAVSKY, Aaron. The Self-Evaluating Organization. In CHELIMSKY, Eleanor (Ed.). **Program Evaluation: Patterns and Directions**. Washington, DC: American Society for Public Administration, 1985, p. 246–265. [Republicação de artigo in **Public Administration Review** Vol. 2, n. 5 (1972), p. 509–520.

WILLIAMS, Walter. Implementation Analysis and Assessment. In WILLIAMS, Walter; ELMORE, Richard F. (Eds.). **Social Program Implementation**. New York: Academic Press, 1976, p. 267–292.

WILSON, James Q. **The Investigators: Managing FBI and Narcotics Agents**. New York: Basic Books, 1978.

WILSON, James Q. **Bureaucracy: What Government Agencies Do and Why They Do It**. New York: Basic Books, 1989.

WINTER, Søren. Integrating Implementation Research. In PALUMBO, Dennis J.; CALISTA, Donald J. (Eds.). **Implementation and the Policy Process: Opening Up the Black Box**. London: Greenwood Press, 1990, p. 19–38.

WOHLFARTH, Janet et. al. (Eds.) **ÅKA-modellen: En berättelse om hur ESF projektet GEVALIS Vuxna bidragit till förnyat arbetssätt genom integrerad samverkan** (The ÅKA model: A tale of how the ESF-project GEVALIS Vuxna contributed to a renewed working method through integrated coproduction). Skara: amordningsförbundet Västra Skaraborg, Report 2014:2 (2nd edition).

WOODSIDE, Kenneth. Policy Instruments and the Study of Public Policy. **Canadian Journal of Political Science/Revue Canadienne de Science Politique**, Vol. 19, n. 4 (1986), p. 775–794.

XU, Runya; SUN, Qigui; SI, Wei. The Third Wave of Public Administration: The New Public Governance. **Canadian Social Science** Vol. 11, n. 7 (2015), p. 11–21.

YIN, Robert K. Studying the Implementation of Public Programs. In WILLIAMS, Walter. (Ed.). **Studying Implementation: Methodological and Administrative Issues**. Chatham, NJ: Chatham House, 1982, p. 36–72.

YIN, Robert K. **Case Study Research: Design and Methods**. Newbury Park: Sage, 1984.

ZETTERBERG, Hans L. **Social Theory and Social Practice**. New York, 1964.

Periódicos

Evaluation: The International Journal of Theory, Research, and Practice. Publicado por Sage Publications.

Evaluation and Program Planning. Publicado por Elsevier.

Evaluation Practice, revista da American Evaluation Association. Além de artigos, este periódico contém uma área dedicada a debates e uma seção de referências bibliográficas a artigos correntes e relatórios de avaliação, apresentando uma linha sobre cada um. Vem sempre seguida da newsletter “Evaluation Practice News” com depoimentos profissionais e notícias pessoais. Desde 1998 tomou o nome *The American Journal of Evaluation*.

Evaluation Review: A Journal of Applied Social Research. Antes Evaluation Quarterly. Provavelmente o mais importante no campo.

Journal of Policy Analysis and Management. Voz oficial da Association for Public Policy Analysis and Management (APPAM). A associação foi fundada em 1979.

New Directions for Program Evaluation (NDE). Uma publicação oficial da American Evaluation Association e talvez um dos melhores periódicos no campo. NDE publica trabalhos empíricos, metodológicos e teóricos em todos os aspectos da avaliação e campos relacionadas. Áreas substantivas incluem qualquer campo, como política de taxaço, energia, meio ambiente, educaço, treinamento de trabalho, saúde pública. Também inclui tópicos como avaliação de produtos, avaliação de pessoal, análise de políticas públicas e avaliação de tecnologia. Cada número da NDE é devotada a um tópico singular, como contribuiçoes solicitadas, organizadas, revisadas e editadas por um editor convidado.

Policy Sciences: Um periódico internacional dedicado à melhoria da formulação de políticas públicas (Policy Making). Início em 1970. Uma importante revista sobre planejamento, análise de políticas públicas e avaliação.

Policy Studies. Publicado pelo Policy Study Institute, Londres, "a organização independente líder no Reino Unido que realiza estudos de políticas econômicas, industriais e sociais assim como o funcionamento das instituições políticas. O Instituto tem como meta informar as políticas públicas, estabelecendo os fatos."

Public Administration. Publicado pela Blackwell Publishing. Periodicidade trimestral. Artigos sobre administração pública e processos decisórios administrativos.

Public Administration Review. Publicado a cada dois meses pela American Society for Public Administration.

BRASIL

Brazilian Political Science Review (ABCP) A BPSR é publicada pela Associação Brasileira de Ciência Política, revista de ciência política e relações internacionais publicada em inglês no Brasil. O principal objetivo da BPSR é apoiar a difusão da obra de ciência política de alta qualidade produzida no Brasil e no exterior, contribuindo para a troca de ideias dentro da comunidade internacional de ciência política.

Planejamento e Políticas Públicas (PPP) – Ipea. A Revista PPP, editada pelo Instituto de Pesquisa Econômica Aplicada (Ipea), possui uma história de publicações de reconhecido valor acadêmico no cenário nacional. O objetivo principal do periódico é promover o debate e a circulação de conhecimento em planejamento e políticas públicas, representando de fato o esforço do Instituto no sentido de disseminar pesquisas, avaliações e proposições neste campo.

Perspectivas em Políticas Públicas (PPP). A Revista PPP é um periódico acadêmico, da Faculdade de Políticas Públicas "Tancredo Neves" - FaPP/CBH/Universidade Estadual de Minas Gerais, que publica produções teórico-práticas no campo das políticas públicas e suas interfaces com a gestão. A Revista PPP tem por objetivo disseminar produções de conhecimentos, provenientes de diálogos da academia com as mais diversas vozes, discursos e experiências sociais.

Política Hoje. UFPe. A Revista Política Hoje é uma publicação eletrônica semestral de artigos inéditos nas áreas de Ciência Política e Relações Internacionais. A revista é dividida em quatro seções: a) **dossiê**: uma seção contendo textos sobre temas específicos, escolhidos a cada edição; b) **artigos**: uma seção dedicada a artigos diversos; c) **resenhas**: seção onde os autores apresentam um resumo bibliográfico de obras recentes; d) **fórum**: seção designada para permitir o diálogo com artigos já publicados na Política Hoje.

Revista de Administração Pública (RAP). FGV-EBAPE. Criada em 1967, a RAP é um dos mais importantes e influentes periódicos da área de Administração Pública. A RAP concentra esforços para servir de instrumento indispensável ao aperfeiçoamento continuado de pesquisadores, professores, gestores e atores sociais e políticos comprometidos com a efetividade e a equidade da ação pública.

Revista de Administração. Universidade de São Paulo. A Revista de Administração tem por objetivo publicar artigos conceituais, práticos e de pesquisa que agreguem valor ao trabalho de acadêmicos e praticantes de Administração, missão que vem cumprindo há mais de 30 anos. Seus leitores são professores, alunos de graduação e pós-graduação, consultores, empresários e profissionais de empresas públicas e privadas.

Revista Ensaio: Avaliação e Políticas Públicas em Educação. Ensaio é uma publicação trimestral da Fundação Cesgranrio, do Rio de Janeiro. Com tiragem de 1.500 exemplares, circula em caráter nacional e internacional, com distribuição e acesso, via internet, gratuita. Ensaio é um veículo de divulgação de pesquisas, levantamentos, estudos, discussões e outros trabalhos críticos no campo da educação, concentrando-se nas questões de avaliação e políticas públicas e enfatizando as experiências e perspectivas brasileiras. Ensaio é pluralista, do ponto de vista das ideias e das escolas de pensamento; interdisciplinar, do ponto de vista das preocupações e metodologias empregadas por seus colaboradores, acolhendo ou solicitando trabalhos sobre uma larga multiplicidade de temas. Os artigos assinados refletem as opiniões de seus autores e não as da Editoria ou do Conselho da revista Ensaio, nem da Fundação Cesgranrio.

Revista Brasileira de Ciência Política (UnB) - publicada pelo Instituto de Ciência Política da Universidade de Brasília, divulga estudos sobre o fenômeno da política- que partam de diferentes perspectivas teóricas e metodológicas e que tenham em comum o compromisso com o rigor metodológico e a capacidade de estimular a pesquisa na área.

Revista Brasileira de Estudos de Defesa (RBED) é um periódico acadêmico semestral que publica artigos científicos, ensaios e resenhas relacionados à área de defesa e segurança internacional. A RBED é um periódico mantido pela **Associação Brasileira de Estudos de Defesa (ABED)**.

Revista Brasileira de Estudos Estratégicos (Inest-UFF). Publicada pelo Instituto de Estudos Estratégicos da Universidade Federal Fluminense, visa a divulgação dos resultados dos estudos e das pesquisas dos pesquisadores, brasileiros e estrangeiros na área dos Estudos Estratégicos. A *REST* tem compromisso inarredável com o desenvolvimento do pensamento estratégico brasileiro. Guarda, por isso, especial atenção para com os trabalhos que possam contribuir para o avanço de metodologias e conceitos inovadores na área como o estudo, a análise e a avaliação de políticas de defesa nacional e de segurança internacional.

Revista Brasileira de Monitoramento e Avaliação. Publicada pela Rede Brasileira de Monitoramento e Avaliação, a RBMA tem o objetivo de proporcionar uma visão plural, integrada e multidisciplinar da área de monitoramento e avaliação em políticas públicas, programas e projetos sociais no Brasil e no mundo. Reunindo artigos e contribuições conceituais, metodológicos e empíricos em análises institucionais, pesquisas quantitativas, qualitativas, quasi-experimentais, estudos de caso, resenhas e registros de memória institucional (ensaios e entrevistas).

Revista Brasileira de Políticas Públicas / Brazilian Journal of Public Policy (Uniceub). Periódico acadêmico da área jurídica que tem como finalidade constituir instrumento de veiculação semestral de trabalhos científicos e doutrinários que abordem questões jurídicas da contemporaneidade e, ainda, aspectos da interação entre Direito e Políticas Públicas. Direciona, portanto, seu objeto de interesse a questões referentes a governabilidade, integração, participação cidadã, desenvolvimento e outros temas envolvendo o Estado, a Sociedade e o Direito.

Revista Brasileira de Políticas Públicas e Internacionais (RPPI). Periódico científico vinculado ao Programa de Pós-Graduação em Gestão Pública e Cooperação Internacional (PGPCI) da Universidade Federal da Paraíba (UFPB).

Revista Direitos Sociais e Políticas Públicas (UNIFAFIBE) Revista do curso de Direito do Centro Universitário UNIFAFIBE – Bebedouro/SP tem como compromisso a difusão de pesquisas de docentes e pesquisadores de sua instituição, bem como de demais autores nacionais e internacionais, colaborando com a interação entre doutrina e jurisprudência, sobretudo no que concerne à temática direitos sociais e políticas públicas, além de suas áreas afins.

Revista Gestão & Políticas Públicas - Destina-se à publicação de artigos originais sobre temas atuais em gestão das políticas públicas, preferencialmente pautados em abordagem interdisciplinar. É um periódico eletrônico semestral da Escola de Artes, Ciências e Humanidades da Universidade de São Paulo (EACH-USP), em associação com a Fundação do Desenvolvimento Administrativo (FUNDAP).

Revista Parcerias Estratégicas Publicação do CGEE-MCTI . A revista "Parcerias Estratégicas" contribui para a divulgação e promoção de debates sobre temas de interesse estratégico para o desenvolvimento científico e tecnológico nacional. Em edições recentes, a revista destacou os estudos sobre manipulações genéticas, bioética, arranjos locais de produção, fundos setoriais, entre outros temas. A atividade de prospecção e avaliação em Ciência e Tecnologia tem espaço cativo na publicação, dada sua afinidade e relevância para as atividades desenvolvidas no Centro. Na seção memória, procura-se resgatar a história da C&T no Brasil e no mundo.

Revista de Pesquisa em Políticas Públicas (RP3). Publicação do Centro de Estudos Avançados de Governo e Administração Pública – Universidade de Brasília (CEAG-UnB), a RP3 surge como um espaço para reflexão conceitual e teórica sobre a área de Políticas Públicas e Administração Pública no Brasil.

Revista de Políticas Públicas (Universidade Federal do Maranhão). Publicação acadêmica do Programa de Pós-Graduação em Políticas Públicas da UFMA, de periodicidade semestral, destinada a publicar trabalhos científicos relevantes para o avanço teórico-prático das Políticas Públicas. Encontra-se em circulação desde 1995. Online e impresso.

Revista do Serviço Público. Publicada pela ENAP destina-se a disseminar conhecimentos e estimular a reflexão e o debate, apoiando o desenvolvimento dos servidores, o seu compromisso com a cidadania e a consolidação de uma comunidade de praticantes, especialistas e interessados nos temas de políticas públicas e gestão governamental.

Associações Profissionais

American Evaluation Association (AEA). Uma associação internacional de avaliadores práticos e profissionais. Os membros trabalham em universidades, institutos de pesquisa, agências governamentais e firmas de consultorias. *American Journal of Evaluation* (antes denominado *Evaluation Practice*) e *New Directions for Program Evaluation* são revistas publicadas trimestralmente. Organiza uma conferência anual com inúmeros painéis e distribui prêmios e reconhecimentos.

Association for Public Policy Analysis and Management (APPAM) é uma organização para as Escolas de Políticas Públicas dos Estados Unidos. Conferência anual. *Journal of Policy Analysis and Management* é publicado quatro vezes por ano.

Canadian Evaluation Association. Similar ao estadodunidense. Conferência anual. Publica o *Canadian Journal of Evaluation*.

European Evaluation Society. Formada em 1994. Publica uma *Newsletter*. Organiza uma conferência bianual. Membros têm acesso à revista *Evaluation*.

Policy Studies Organization. Uma das principais associações profissionais dos Estados Unidos, que publica dois importantes periódicos -*Policy Studies Journal* e *Review of Policy Research*.

Swedish Evaluation Society. Conferência a cada 18 meses.

BRASIL (associações profissionais e acadêmicas relacionadas com políticas públicas)

RBM&A – Rede Brasileira de Monitoramento e Avaliação. Continuidade da primeira iniciativa brasileira, a REBRAMA. Foi (re)fundada em novembro de 2008, com o apoio da Fundação João Pinheiro (FJP), do Banco Interamericano de Desenvolvimento, e do Banco Mundial. Os principais objetivos da Rede são: a) -Promover o uso de práticas de Monitoramento e Avaliação; b) Promover trocas de experiência e de instrumentos de trabalho; c) Facilitar a identificação das instituições trabalhando neste tema no Brasil; d) Formar uma ‘Comunidade de prática.’

ANESP – Associação Nacional dos Especialistas em Políticas Públicas e Gestão Governamental A carreira do EPPGG (Especialista em Políticas Públicas e Gestão Governamental) destina-se à execução de atividade de formulação, implementação e avaliação de políticas públicas, bem assim de direção e assessoramento em escalões superiores da Administração Direta e Autárquica (Lei 7834 12/08/1989, art 1º). Na prática, os EPPGGs são os servidores responsáveis por assessorar, dirigir, coordenar, planejar e gerenciar atividades de formulação, implementação e avaliação de políticas públicas, em diferentes graus de complexidade, responsabilidade e autonomia. O caráter estratégico da atuação dos gestores faz com que eles enfrentem problemas complexos no dia-a-dia, que exigem habilidades de negociação e decisão; conhecimentos sobre áreas como administração pública, políticas públicas, economia e direito; e aprofundamento em temas específicos à área de governo em que trabalham.

ABCP – Associação Brasileira de Ciência Política. Fundada em 1986, a ABCP é uma entidade sem fins lucrativos, que reúne centenas de acadêmicos - pesquisadores, professores e estudantes - e profissionais da área de Ciência Política e outras correlatas. Ao longo de sua história, a ABCP esteve envolvida, através da atuação direta de seus associados, na construção da memória da Ciência Política brasileira. Nos últimos 30 anos, a associação promoveu parcerias com diversas fundações nacionais e internacionais, institutos de pesquisa, agências de fomento, universidades, e outras entidades públicas e privadas, cujos recursos materiais, imateriais e humanos foram de vital importância para a criação de centenas de eventos acadêmicos e outras iniciativas, que acabaram por promover a expansão da área através dos programas de pós-graduação e, mais recentemente, dos cursos de bacharelado em Ciência Política e Relações Internacionais.

ABED – Associação Brasileira de Estudos de Defesa. Criada em 2005, a ABED é uma associação acadêmica que tem por finalidade congregar pesquisadores que desenvolvam estudos e pesquisas sobre defesa nacional, segurança nacional e internacional, guerra e paz, relações entre Forças Armadas e sociedade, ciência e tecnologia no âmbito da defesa nacional e questões militares em geral, assim como demais assuntos de natureza estratégica.

ANPAD – Associação Nacional de Pós-Graduação e Pesquisas em Administração. A ANPAD desenvolve um consistente trabalho na promoção do ensino, da pesquisa e na produção de conhecimento dentro do campo das ciências administrativas, contábeis e afins no Brasil. Congrega programas de pós-graduação *stricto sensu*, representando os interesses das instituições filiadas junto à opinião pública e atuando como órgão articulador dos interesses dos programas perante a comunidade científica e os órgãos governamentais responsáveis pela gestão da educação e desenvolvimento científico e tecnológico em nosso país. Criada em 1976, a partir da iniciativa dos oito programas de pós-graduação então existentes no Brasil, a ANPAD é hoje o principal órgão de interação entre programas associados, grupos de pesquisa da área e a comunidade internacional. Aliado a sua sólida atuação, o crescimento considerável de cursos de pós-graduação ofertados fez com que a Associação comemorasse seus 40 anos de atividades reunindo mais de 100 programas associados em uma comunidade acadêmica internacionalmente reconhecida.

ANPOCS – Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais - Fundada em 1977, a Anpocs, entidade de direito privado sem fins lucrativos, reúne mais de uma centena de centros de pós-graduação e de pesquisa em antropologia, ciência política, relações internacionais, sociologia, de todo o Brasil. Diferentemente de outras associações científicas, a Anpocs é composta por sócios institucionais e não por pesquisadores individuais. Nossos afiliados contam com mais de 1200 professores universitários e pesquisadores, profissionais de alto nível, além de milhares de estudantes de mestrado e doutorado em centros localizados de norte a sul do país.

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. O INEP é uma autarquia federal vinculada ao Ministério da Educação (MEC) cuja missão é subsidiar a formulação de políticas educacionais dos diferentes níveis de governo com intuito de contribuir para o

desenvolvimento econômico e social do país. Além de avaliar a educação básica e educação superior, assumiu a avaliação integrada da pós-graduação. O Inep publica o *Glossário dos Instrumentos de Avaliação Externa* (2018, 2ª edição).

ABAVE – Associação Brasileira de Avaliação Educacional. "A ABAVE é um espaço plural e democrático para o intercâmbio de experiências entre os acadêmicos e os implementadores da avaliação educacional. O que justifica a sua criação é o crescimento, nos últimos anos, do número de pessoas que lidam com a avaliação e que começam a vê-la como seu campo de atuação profissional."

ABOP – Associação Brasileira de Orçamento Público. A ABOP é uma entidade civil, sem fins lucrativos, fundada em 04 de dezembro de 1974, com domicílio jurídico em Brasília-DF, regionais estabelecidas em diversos estados da Federação e filiada à Asociación Internacional de Presupuesto Público - ASIP, com os seguintes propósitos: a) Atuar na criação, divulgação, aplicação e desenvolvimento de novas técnicas, conceitos, procedimentos e normas no campo orçamentário e correlatos; b) Reciclar e capacitar profissionais de planejamento e orçamento público e áreas afins, de todos os poderes e níveis de governo; c) Promover eventos de aperfeiçoamento das técnicas orçamentárias, disponibilizando aos associados o instrumental adequado para análise e solução de questões específicas; d) Estimular a realização de estudos e pesquisas no campo teórico, conceitual e prático sobre problemas relativos ao orçamento; e) Propor e desenvolver formas de participação da sociedade nas decisões incidentes sobre o planejamento, execução orçamentária, gerenciamento e fiscalização das políticas públicas; f) Servir de fonte de consulta permanente, a nível técnico de informação, aos organismos nacionais e internacionais.

CGEE – Centro de Gestão e Estudos Estratégicos (MCTI). É uma associação privada sem fins lucrativos, fundada em setembro de 2001, no âmbito da 2ª Conferência Nacional de Ciência e Tecnologia. No encontro, em que se elegeu a inovação como tema crucial para o desenvolvimento científico e tecnológico do país, 273 pesquisadores e especialistas, vinculados a aproximadamente 100 instituições de 22 estados, endossaram a proposta do então Ministério da Ciência e Tecnologia (MCT) de criar um órgão que subsidiasse a Política Nacional de Ciência, Tecnologia e Inovação (CT&I) e as decisões de longo prazo dos setores público e privado em temas relacionados. Em 2002, foi classificado como organização social, por meio do Decreto nº 4.078, de 9 de janeiro daquele ano.

Apoio

Produção Científica

INSTITUTO DE ESTUDOS ESTRATÉGICOS
UNIVERSIDADE FEDERAL FLUMINENSE

Produção e Divulgação

EDITORA LUZES
COMUNICAÇÃO, ARTE & CULTURA
Tel.: (21) 2447-4336
E-mail: editoraluzes@editoraluzes.com.br
Site: www.editoraluzes.com.br

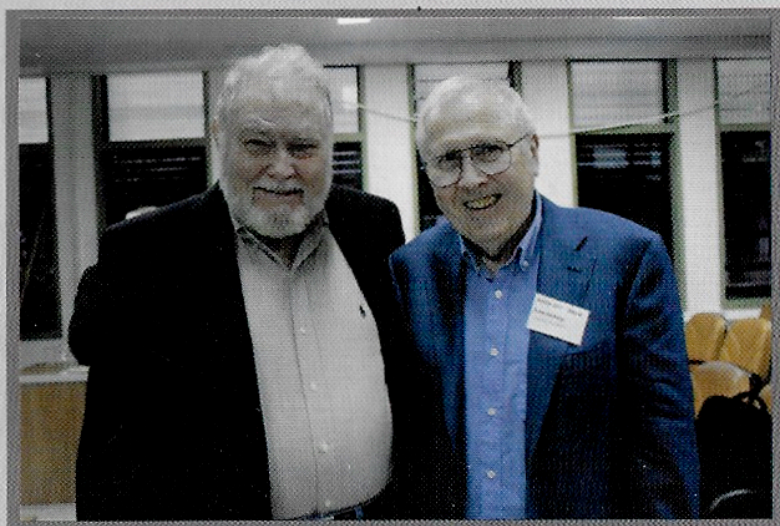
O INEST E SUA PRODUÇÃO CIENTÍFICA

O INEST, ao longo de sua trajetória como instituição aos Estudos Estratégicos, vem se consolidando com inúmeras pesquisas em favor do conhecimento.

Essa participação se estende na área de defesa, das relações internacionais, priorizando esses temas em publicações que contêm o saber de seus autores sejam em artigos, resenhas de livros, textos extraídos de teses, dissertações e monografias em geral. Além da edição de sua Revista, estabelecendo um forte dossiê de fontes primárias, avança pela área da publicação de livros de seus pesquisadores.

Esta obra, por exemplo, "Avaliação de Políticas Públicas e Programas Governamentais", de autoria dos professores Evert Vedung e Luiz Pedone, preenche uma lacuna que instiga, cada vez mais, os pesquisadores, enriquecendo as estantes científicas.

Nestas páginas vamos encontrar um profundo manancial de conhecimentos que, certamente, contribuirá para a elevada diversificação dos assuntos aqui explorados, provocando a necessária discussão de um tema hoje, tão excitante, pela sua complexa natureza. Assim, ampliada pela diversificada exploração e qualidade dos conteúdos, vem beneficiar os pesquisadores do tema.



Luiz Pedone (E) e Evert Vedung

EVERT VEDUNG, professor emérito de ciência política e políticas públicas na Universidade de Uppsala, Suécia, é professor visitante de avaliação na Universidade Federal Fluminense e Universidade de Brasília e instituições como o Tribunal de Contas da União, ENAP, Senado Federal e IPEA.

LUIZ PEDONE, professor associado de políticas públicas e relações internacionais na Universidade Federal Fluminense. Foi professor de políticas públicas e chefe do departamento de ciência política na Universidade de Brasília (1976-2003). Ministra cursos de avaliação com o professor Vedung desde 2013 na UFF.

«O livro *Avaliação de Políticas Públicas e Programas Governamentais* é um livro essencial pois analisa temas importantes da avaliação de políticas públicas de forma sistemática e atualizada.»

Paulo Calmon, IPOL, UnB

Produção



Comunicação, Arte & Cultura

ISBN 97865992207-6-0

