

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE CIÊNCIAS ECONÔMICAS
Departamento de Ciências Administrativas
Centro de Pesquisas e Pós-Graduação em Administração

Cinthia Helena de Oliveira Bechelaine

AS AVALIAÇÕES VÃO PARA A GAVETA?

Um Estudo da Utilização de Avaliações de Impacto no Governo de Minas Gerais

Belo Horizonte
2013

Cinthia Helena de Oliveira Bechelaine

AS AVALIAÇÕES VÃO PARA A GAVETA?

Um Estudo da Utilização de Avaliações de Impacto no Governo de Minas Gerais

Dissertação apresentado ao Curso de Mestrado em Administração do Centro de Pós-Graduação e Pesquisas em Administração, da Faculdade de Ciências Econômicas, da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Administração.

Área de Concentração – Estudos Organizacionais e Sociedade.

Orientador – Prof. Ivan Beck Ckagnazaroff, PhD.

Belo Horizonte

2013

Ficha catalográfica

B391a
2013

Bechelaine, Cinthia Helena de Oliveira, 1987-
As avaliações vão para a gaveta? [manuscrito] : um estudo da utilização de avaliação de impacto no governo de Minas Gerais / Cinthia Helena de Oliveira Bechelaine / 2013.
241 f.: il.

Orientador: Ivan Beck Ckagnazaroff.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Centro de Pós-Graduação e Pesquisas em Administração.
Inclui bibliografia (p. 233-239) e apêndice.

1. Políticas públicas – Minas Gerais – Teses. 2. Políticas públicas – avaliação – Teses. I. Universidade Federal de Minas Gerais. Centro de Pós-Graduação e Pesquisas em Administração. II. Ckagnazaroff, Ivan Beck. III. Título.

CDD: 361.6198151

Elaborada pela Biblioteca da FACE/UFMG. – NMM/038/2013

Universidade Federal de Minas Gerais
Faculdade de Ciências Econômicas
Departamento de Ciências Administrativas
Centro de Pós-Graduação e Pesquisas em Administração

ATA DA DEFESA DE DISSERTAÇÃO DE MESTRADO EM ADMINISTRAÇÃO da Senhora CINTHIA HELENA DE OLIVEIRA BECHELAINE, REGISTRO Nº 511/2013. No dia 18 de abril de 2013, às 09:30 horas, reuniu-se na Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais - UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Centro de Pós-Graduação e Pesquisas em Administração do CEPEAD, em 22 de março de 2013, para julgar o trabalho final intitulado "As Avaliações vão para a Gaveta? Um Estudo da Utilização de Avaliações de Impacto no Governo de Minas Gerais", requisito para a obtenção do Grau de Mestre em Administração, área de concentração: Administração. Abrindo a sessão, o Senhor Presidente da Comissão, Prof. Dr. Ivan Beck Ckagnazaroff, após dar conhecimento aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a argüição pelos examinadores com a respectiva defesa da candidata. Logo após, a Comissão se reuniu sem a presença da candidata e do público, para julgamento e expedição do seguinte resultado final:

(X) APROVAÇÃO;

( ) APROVAÇÃO CONDICIONADA A SATISFAÇÃO DAS EXIGÊNCIAS CONSTANTES NO VERSO DESTA FOLHA, NO PRAZO FIXADO PELA BANCA EXAMINADORA (NÃO SUPERIOR A 90 NOVENTA DIAS);

( ) REPROVAÇÃO.

O resultado final foi comunicado publicamente à candidata pelo Senhor Presidente da Comissão. Nada mais havendo a tratar, o Senhor Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 18 de abril de 2013.

NOMES ASSINATURAS

Prof. Dr. Ivan Beck Ckagnazaroff.....
ORIENTADOR (CEPEAD/UFMG)

Prof. Dr. Afonso Henriques Borges Ferreira.....
(ECM/UFMG)

Prof. Dr. Eduardo Luiz Gonçalves Rios Neto.....
(CEDEPLAR/ UFMG)

Prof. Dr. Carlos Aurélio Pimenta de Faria.....
(PPGCS/PUC MINAS)

A minha mãe, exemplo de dedicação incondicional.

AGRADECIMENTOS

Gostaria de agradecer a todos que direta ou indiretamente contribuíram para o cumprimento de mais esta etapa.

Em primeiro a minha família, pela dedicação, compreensão quanto às ausências e pela paciência com todas as ansiedades advindas do mestrado.

Não poderia deixar de agradecer ao professor e orientador Ivan Beck Ckagnazaroff, que me conduziu com paciência e afeição durante esta pesquisa. Obrigada pelos debates, cafés e pelos conselhos sinceros.

Agradeço também a todos os profissionais e colegas de trabalho do governo de Minas Gerais, a quem devo a inspiração ao tema e debates sobre o assunto. Em especial à Vanda, Michel, Diana que acompanharam as etapas desse processo desde a candidatura ao mestrado.

Por fim, agradeço a disponibilidade de todos aqueles que contribuíram para o estudo e participaram das entrevistas.

"The most important purpose of evaluation is not to prove but to improve".

Daniel L. Stufflebeam

RESUMO

Apesar de as avaliações de políticas públicas constituírem subsídios para que a tomada de decisões seja mais racional, motivada por um conjunto de informações confiáveis sobre a implementação e o desempenho de programas governamentais, nem sempre são utilizadas como deveriam. Autores internacionais observam que os estudos avaliatórios acabam sendo mal utilizados, esquecidos ou engavetados. No Brasil e nos governos estaduais, apesar do crescente interesse em avaliações de programas governamentais, é possível observar certa escassez de estudos empíricos nacionais que permitam verificar a utilização de tais avaliações como instrumento de correção de rumos e aprimoramento dos serviços públicos. Nesse contexto, a presente pesquisa tem por objetivo analisar a dinâmica da utilização dos resultados das avaliações de projetos estruturadores realizadas pelo governo do estado de Minas Gerais a desde o ano de 2007, a partir de uma pesquisa qualitativa que utiliza a estratégia do estudo de caso e emprega as técnicas de observação direta, análise documental e, principalmente, entrevistas semiestruturadas com diferentes *stakeholders* envolvidos nessas avaliações. Com vistas a ilustrar e problematizar a elaboração teórica, o problema da utilização das avaliações é estudado nos casos de avaliações de quatro avaliações de projetos estruturadores do governo mineiro: Viva Vida, Gestão Integrada de Ações e Informações de Defesa Social, Minas Sem Fome e Acelerar para Vencer. Em cada um dos casos é remontado o histórico da avaliação, desde a concepção da avaliação e expectativa inicial, até a conclusão da mesma, com alusão ao tipo de utilização ocorrido, pela percepção dos entrevistados. Conclui-se que o estado de Minas Gerais apresenta um sistema de avaliação em construção, ainda recente, na administração pública brasileira e que requer aprimoramento em seu arranjo institucional. A incapacidade de acomodar o pluralismo de atores envolvidos nos programas públicos, a desconsideração do contexto, das circunstâncias e do propósito que envolve a implantação de programas são alguns dos fatores responsáveis pela não utilização dos resultados dos estudos avaliatórios.

Palavras-chave: políticas públicas, avaliação de programas, utilização de avaliações.

ABSTRACT

Although program evaluations constitute subsidies to make decision process more rational, motivated by a number of reliable information on the implementation and performance of government programs, they are not always used as they should. As international authors note, evaluative studies end up being misused, forgotten or even shelved. In Brazil and in the Brazilian state governments, despite growing interest in program evaluation, it is possible to observe shortage of empirical national studies that verify the use of such assessments as a tool for public services course corrections and improvement. In this context, this research aims to analyze the dynamics of the use of evaluation results of structuring projects undertaken since 2007 by the State Government of Minas Gerais. It is a qualitative research strategy that uses the case study and employs the techniques of direct observation, document analysis and, mainly, semi-structured interviews with different stakeholders involved in these assessments. In order to illustrate and discuss the theoretical research, the problem of the evaluation use is studied in cases of evaluations of four structuring projects evaluation: *Viva Vida, Gestão Integrada de Ações e Informações de Defesa Social, Minas Sem Fome* and *Acelerar para Vencer*. In each of the cases are traced back the history of evaluation, since evaluation design and initial expectation, until it completion, with reference to the type of use occurred, using the respondents perception. We conclude that the State of Minas Gerais presents an evaluation system under construction, still fresh in Brazilian public administration and requires improvement in its institutional arrangement. The inability to accommodate the plurality of actors involved in public programs, ignoring the context, circumstances and purpose that involves the deployment of programs are some of the factors responsible for not using the results of evaluative studies.

Key words: public policy, program evaluation, evaluation use.

LISTA DE ILUSTRAÇÕES

1) Lista de figuras

Figura 1 – Relações entre a utilização das avaliações de Alkin e Coyle (1988). 47
Figura 2 – Método de Estudos de Caso Múltiplos 83

2) Lista de quadros

Quadro 1 – Diferenças entre monitoramento, auditoria e avaliação 32
Quadro 2 – Padrões do Joint Committee on Standards for Educattional Evaluation (1994) ... 40
Quadro 3 – Diferentes usos das avaliações 51
Quadro 4 – Fatores relacionados à utilização das avaliações na pesquisa de Cousins e Leithwood (1986) 54
Quadro 5 – Fatores relacionados à utilização das avaliações para Leviton e Hughes (1981) .. 62
Quadro 6 – Fatores relacionados à utilização das avaliações para Johnson et al. (2009) 68
Quadro 7 - Relação dos Programas Avaliados 85
Quadro 8 – Relação dos entrevistados 89

LISTA DE ABREVIATURAS E SIGLAS

AEA	*American Evaluation Association*
ACISP	Áreas de Coordenação Integrada de Segurança Pública
AISP	Áreas Integradas de Segurança Pública
ALMG	Assembleia Legislativa de Minas Gerais
BDMG	Banco de Desenvolvimento de Minas Gerais
BID	Banco Interamenicano de Desenvolvimento
BIRD	Banco Mundial
CBMG	Corpo de Bombeiros Militar
CEDEPLAR	Centro de Desenvolvimento e Planejamento Regional da Universidade Federal de Minas Gerais
CRISP	Centro de Estudos de Criminalidade e Segurança Pública da Universidade Federal de Minas Gerais
CVV	Centros Viva Vida
EMATER	Empresa de Assistência Técnica e Extensão Rural do Estado de Minas Gerais
EpR	Programa Estado para Resultados
FJP	Fundação João Pinheiro
FUNDEP	Fundação de Desenvolvimento da Pesquisa
GAO	*Government Auditing Standards*
IDEB	Índice de Desenvolvimento da Educação Básica
IDH	Índice de Desenvolvimento Humano
IGESP	Integração da Gestão em Segurança Pública
IPEA	Instituto de Pesquisa Econômica Aplicada
JCSEE	*Joint Committe on Standards for Educational Evaluation*
LDO	Lei de Diretrizes Orçamentárias
LOA	Lei do Orçamento Anual
LRF	Lei de Responsabilidade Fiscal
MDS	Ministério do Desenvolvimento Social e Combate à Fome
M&A	Monitoramento e Avaliação
MSF	Minas Sem Fome
NESP	Núcleo de Estudos em Segurança Pública
NPM	*New Public Management*

OMS	Organização Mundial de Saúde
PAV	Projeto Acelerar para Vencer
PCMG	Polícia Civil de Minas Gerais
PEP	Programa de Ensino Médio Profissionalizante
PDR	Plano Diretor da Regionalização
PIP	Programa de Intervenção Pedagógica
PMDI	Plano Mineiro de Desenvolvimento Integrado
PMMG	Polícia Militar de Minas Gerais
PPA	Plano Plurianual
PPAG	Plano Plurianual de Gestão Governamental
PROALFA	Programa de Avaliação da Alfabetização
PROEB	Programa de Avaliação da Rede Pública de Educação Básica
PROETI	Projeto Escola em Tempo Integral
PSF	Programa Saúde da Família
RMBH	Região Metropolitana de Belo Horizonte
SCGERAES	Superintendência Geral de Gestão Estratégica de Recursos e Ações do Estado
SCPPO	Superintendência Central de Planejamento e Programação Orçamentária
SEAPA	Secretaria de Agricultura, Pecuária e Abastecimento de Minas Gerais
SEDESE	Secretaria de Estado de Desenvolvimento Social
SEDS	Secretaria de Estado de Defesa Social de Minas Gerais
SEE	Secretaria de Estado de Educação de Minas Gerais
SEPLAG	Secretaria de Estado de Planejamento e Gestão de Minas Gerais
SES	Secretaria de Saúde do Estado de Minas Gerais
SIDS	Sistema Integrado de Defesa Social
SIMAVE	Sistema Mineiro de Avaliação da Educação Pública
SUMIM	Superintendência Central de Modernização Institucional
SUS	Sistema Único de Saúde
UBS	Unidade Básica de Saúde

Sumário

1 INTRODUÇÃO 15
1.1 Problema de pesquisa e sua relevância 19
1.2 Objetivos de pesquisa 23

2 AS AVALIAÇÕES DE PROGRAMAS GOVERNAMENTAIS 25
2.1 Avaliações: de onde vêm? 25
2.2 Conceitos e abordagens 28
2.3.1 Tipos de avaliação 32
2.4 Como avaliar as avaliações? 36

3 A UTILIZAÇÃO DOS RESULTADOS DAS AVALIAÇÕES 43
3.1 Discutindo a utilização 43
3.2 As dimensões do uso 48
3.3 As vaiáveis que influenciam a utilização das avaliações 52
3.3.1 Cousins e Leithwood (1986): a importância do contexto 53
3.3.2 Shulha e Cousins (1997): o papel da política e das estruturas e processos organizacionais 63
3.3.3 Johnson *et al.* (2009): a importância do envolvimento dos *stakeholders* 66

4 AS AVALIAÇÕES DE PROGRAMAS NO SETOR PÚBLICO 70
4.1 Antecedentes: reformas da administração pública e o contexto brasileiro 70
4.2 O Choque de Gestão em Minas Gerais 74
4.3 A gestão por resultados em Minas Gerais 78

5 METODOLOGIA 80
5.1 Os estudos de casos múltiplos 80
5.2 Seleção dos casos 84
5.3 Coleta e análise de dados 86
5.4 Dificuldades e limitações do estudo 90

6 AS AVALIAÇÕES DOS PROJETOS ESTRUTURADORES NO GOVERNO DE MINAS GERAIS 93

6.1 Bases do movimento de avaliação e atores envolvidos 93
6.2 Processo geral de implementação e execução das avaliações 103

7 AVALIAÇÃO DO PROJETO VIVA VIDA 114
7.1 Contexto geral da avaliação 114
7.2 Implementação do processo de avaliação e envolvimento dos atores 120
7.3 Resultados e processamento de informações 128
7.4 Utilização da avaliação do ponto de vista dos *stakeholders* 133

8 AVALIAÇÃO DO PROJETO GESTÃO INTEGRADA DE AÇÕES E INFORMAÇÃO DE DEFESA SOCIAL 140
8.1 Contexto geral da avaliação 140
8.2 Implementação do processo de avaliação e envolvimento dos atores 147
8.3 Resultados e processamento de informações 155
8.4 Utilização da avaliação do ponto de vista dos *stakeholders* 160

9 AVALIAÇÃO DO PROJETO MINAS SEM FOME 167
9.1 Contexto geral da avaliação 167
9.2 Implementação do processo de avaliação e envolvimentos dos atores 173
9.3 Resultados e processamento de informações 181
9.4 Utilização da avaliação do ponto de vista dos *stakeholders* 184

10 AVALIAÇÃO DO PROJETO ACELERAÇÃO DA APRENDIZAGEM NO NORTE DE MINAS, JEQUITINHONHA, MUCURI E RIO DOCE 190
10.1 Contexto geral da avaliação 190
10.2 Implementação do processo de avaliação e envolvimento dos atores 199
10.3 Apresentação dos resultados e processamento de informações 209
10.4 Utilização da avaliação do ponto de vista dos *stakeholders* 215

11 CONSIDERAÇÕES FINAIS 224

REFERÊNCIAS 233

APÊNDICE 240

1 INTRODUÇÃO

O conceito de avaliações de políticas públicas, assim como o de planejamento, surge a partir das transformações no papal do Estado à adoção de políticas sociais e a consequente necessidade de analisar os custos e as vantagens de sua intervenção (ROSSI; FREEMAN, 1993). Mais recentemente, no âmbito do processo de mudança da administração pública gerencial, o número de governos que trabalha para aprimorar a performance da sua atuação, via criação de sistemas para medir e compreender seu desempenho, cresce consideravelmente. Esses sistemas de avaliação são usados para medir a quantidade, a qualidade e a segmentação dos produtos e serviços que o Estado oferece, de forma a facilitar a compreensão das causas do desempenho de programas governamentais.

Segundo a definição de Ala-Harja e Helgason (2000), por programa governamental entende-se um conjunto de atividades organizadas dentro de um cronograma e orçamento determinado do que se dispõe para a implementação de políticas, ou seja, para a criação de condições que permitam o alcance de metas políticas desejáveis. Em algum momento governos e cidadãos irão necessitar de distinguir os programas governamentais com capacidade transformadora daqueles que são inefetivos, revisar as políticas em andamento ou, até mesmo, lançar novos programas no intuito de alcançar os resultados desejados. Para fazer isso, eles precisam de informações contidas em avaliações (ROSSI; FREEMAN, 1993).

No caso de uma política ou programa público, o que a avaliação faz é utilizar métodos científicos para que o processo de julgamento se torne sistemático e confiável (WEISS, 1998). Para isso, os avaliadores podem utilizar questionários, coletar evidências das pessoas envolvidas nos programas governamentais, traduzir dados em indicadores quantitativos e, assim, por meio de várias maneiras, construir conclusões sobre os programas. Essas conclusões, seja sobre os processos dos programas, seja sobre sua efetividade, devem ser endereçadas àqueles responsáveis por patrocinar, gerenciar e executar os programas ou mesmo participar de suas atividades.

Na visão de Costa e Castanhar (2003), a avaliação é um instrumento de mudança poderoso que não deve apenas demonstrar os problemas ou propor soluções para os problemas detectados, mas produzir mudanças apropriadas que conduzam ao cumprimento de padrões

mínimos de qualidade em diferentes serviços. Nesse sentido, o objetivo não é outro senão a identificação, avaliação e correção de situações potencialmente melhoráveis.

A prática de avaliar as políticas públicas ganhou impulso a partir dos vultosos investimentos em políticas sociais realizados, principalmente, por governos norte-americanos na década de 1960 (WEISS, 1998). Como consequência, a partir desse período, o campo de estudos na área de avaliação de políticas prosperou, determinando o aparecimento de atores de diferentes campos que buscavam encontrar novos e melhores modelos e métodos mais adequados para se avaliarem as políticas sociais.

O processo de institucionalização da avaliação na América Latina e no Brasil ocorreu tardiamente, aliada ao conjunto de mudanças instituídas pelo movimento de reforma do Estado (FARIA, 2005). A avaliação dos programas no Brasil ganha força praticamente nos anos 1990, motivada por fatores como a crise de financiamento do setor público e as iniciativas de organismos internacionais de financiamento, tais como o Banco Mundial (BIRD) e o Banco Interamericano de Desenvolvimento (BID), no sentido de instrumentalizar países em desenvolvimento para gestão pública mais eficiente e efetiva (RIBEIRO, 2009).

Seguindo o movimento nacional de reforma do Estado, começa a surgir um protagonismo estadual em torno de princípios da administração gerencial e da cultura de resultados. Esse caso pode ser ilustrado pelo Estado de Minas Gerais, cujo planejamento estratégico, a gestão pública por resultados e a cultura de avaliações tornaram-se aspectos marcantes do processo de implementação da reforma administrativa (CORRÊA, 2007).

No entanto, embora tenham ganhado importância considerável e pareçam *a solução de todos os problemas*, as avaliações de programas governamentais abarcam uma série de definições, diferentes metodologias existentes, objetivos e abordagens diversas. Além dos seus inúmeros conceitos, as avaliações normalmente são realizadas com base na premissa de que os resultados serão, direta e automaticamente, utilizados para promover as melhorias necessárias – o que nem sempre ocorre. De acordo com Ospina (2002), a existência de avaliações por si só não garante o melhoramento da gestão pública, vez que a viabilidade desse sistema para promover uma cultura de verificação de resultados e fortalecimento de instituições depende, em grande parte, da utilização da informação gerada.

Vários autores (GUBA; LINCOLN, 1989; SHULHA; COUSINS, 1997; WEISS, 1998; PATTON, 2008) declaram que as informações provindas da avaliação são utilizadas de forma insuficiente. Shulha e Cousins (1997) reiteram que a questão da utilização das avaliações é uma preocupação relevante, talvez a maior fonte de frustração entre os avaliadores. Ala-Harja e Helgason (2000) argumentam que o uso apropriado dos resultados das avaliações não constitui tarefa fácil. Segundo os autores, as avaliações representam, historicamente, um universo de promessas que não chegam a se realizar.

Weiss (1998) argumenta, com base em evidências de estudos de avaliação, que as decisões relativas à implementação de melhorias no âmbito dos programas, tanto no nível gerencial, quanto como na esfera política, pouco são influenciadas por resultados de avaliações. Silva e Melo (2000) apontam que as vicissitudes e obstáculos da implementação de avaliações como instrumento para correção de rota têm natureza diversificada como, por exemplo, aspectos relacionados à capacidade institucional, problemas de natureza política e resistência de grupos ou setores negativamente afetados.

Nota-se que, embora o campo das avaliações de programas tenha experimentado um interesse súbito nas últimas décadas, aliciando profissionais de diferentes áreas de estudo e atraindo, cada vez mais, um volume maior de recursos governamentais, no Brasil, pouca atenção tem sido dada à devida apropriação dos resultados das avaliações por aqueles interessados em suas informações. Mesmo que os estudos avaliatórios cresçam e se desenvolvam cada vez mais, com métodos sofisticados, com dados confiáveis sobre a realidade na qual se intervém, alocando esforços físicos, intelectuais e financeiros, o que os autores da área observam é que o *modus operandi* da intervenção pouco se altera, poucas (ou até nenhuma) das recomendações das avaliações são incorporadas na definição de novos rumos para os programas.

Nesse sentido, dada essa preocupação, a presente pesquisa tem seu foco na utilização dos resultados das avaliações, buscando evidenciar os principais fatores que favorecem ou bloqueiam o devido uso das mesmas de forma a aprimorar os programas governamentais, em especial no governo do estado de Minas Gerais.

Para cumprir esse objetivo, esta dissertação está compreendida em onze capítulos principais, a começar por esta introdução. Nela, além da apresentação do tema a ser pesquisado, são

expostos o problema de pesquisa e sua justificativa, bem como os objetivos, geral e específicos, como será exposto a seguir.

No capítulo seguinte estão os aspectos conceituais do campo das avaliações de programas, histórico, abordagens e diferenças para outros tipos de estudos, quais sejam a pesquisa científica, a auditoria e o monitoramento de políticas. Apresentam-se, também, os principais tipos de avaliações de acordo com o momento em que se avalia, o seu papel ou função, a procedência dos avaliadores e os aspectos do programa que são objeto da avaliação, com foco especial nas avaliações de efetividade. Por fim, pautado na premissa de que todo estudo avaliatório vai ser tendencioso em certa medida, mostra-se, de forma breve, a meta-avaliação, ou seja, a avaliação da avaliação. Nessa oportunidade também serão vistos os padrões do *Joint Committe on Standards for Educational Evaluation* (JCSEE). Procura-se, com esse capítulo, delimitar o tema da pesquisa, ressaltar os autores proeminentes na área e formar as bases para a discussão que será relatada no capítulo seguinte, real foco do trabalho: a utilização dos resultados das avaliações.

No capítulo três, são apresentados os principais autores que se concentraram no tema da utilização. Na primeira seção do capítulo, discute-se o conceito de utilização, má utilização e da não utilização. Na seção seguinte, mostram-se os diferentes usos presentes nas avaliações de programas, ou seja, as formas como as mesmas podem ser utilizadas. Por fim, na última seção do capítulo, enumeram-se os diferentes autores que abordam o *calcanhar de Aquiles* das avaliações: as variáveis que influenciam (ou não) a utilização das avaliações na melhoria de programas públicos. Destacam-se três estudos que contribuíram para a pesquisa sobre as variáveis que influenciam a utilização das avaliações, quais sejam, o estudo de Cousins e Leithwood (1986), o trabalho de Shulha e Cousins (1997) e, por fim, (3) o estudo Johnson *et al.* (2009).

Discutidos o conceito de avaliações e meta-avaliações e indicados os principais referenciais que tratam da utilização das avaliações, é apresentada, no quarto capítulo, uma visão geral dos contextos das avaliações de políticas públicas no Brasil e em Minas Gerais. Dessa forma, no capítulo quarto são mostrados os antecedentes das reformas do setor público nacional que incluíram as avaliações como um dos requisitos para a melhoria da qualidade da provisão de serviços públicos e políticas mais efetivas que culminaram em uma onda de reformas nos governos estaduais. Nessa oportunidade, são relatadas as principais mudanças efetuadas pelo

governo mineiro nas políticas do Choque de Gestão e como a questão da avaliação foi inserida nessa agenda governamental.

No capítulo cinco, explicita-se a metodologia de estudo selecionada para melhor responder o problema de pesquisa, como se deu a coleta e análise de dados e quais foram as principais dificuldades e limitações do estudo. Em linhas gerais, trata-se de uma pesquisa qualitativa que utiliza a estratégia do estudo de caso e emprega as técnicas de observação direta, análise documental e, principalmente, entrevistas semiestruturadas com diferentes *stakeholders* envolvidos nessas avaliações.

Com vistas a ilustrar e problematizar a elaboração teórica, após a explicação da metodologia de pesquisa utilizada, o capítulo seis inaugura a parte exploratória desta pesquisa, qual seja, o estudo de caso da utilização das avaliações empreendidas pelo estado de Minas Gerais a partir de 2007, a partir da visão dos *stakeholders* envolvidos no processo. Em seguida, partindo para os resultados encontrados, os capítulos de sete a dez retratam como se deu a utilização das avaliações dos projetos estruturadores selecionados para a avaliação, sendo um capítulo para cada caso de avaliação. Por fim, no capítulo onze, são apresentadas as considerações finais deste trabalho, seguida pelas referências utilizadas. De início, a seguir, está o problema de pesquisa e sua relevância.

1.1 Problema de pesquisa e sua relevância

De acordo com Cohen e Franco (1993), governos enfrentam, constantemente, situações de aumento das necessidades sociais e diminuição de recursos orçamentários. Ante tal realidade, seria desejável que os governos estudassem as melhores alternativas para a implementação de programas governamentais e para aqueles que já estão em andamento, buscando implementar intervenções com maior benefício social para a população. Nesse sentido, como dito pelos autores, torna-se especialmente importante centrar-se em metodologias para a avaliação de programas governamentais, já que, em uma situação de escassez de recursos e necessidades incrementais, será ainda mais imprescindível a tarefa de comparar, escolher e descartar diferentes formas de intervenção, de forma a garantir certa racionalidade nas ações adotadas.

Atualmente, como relatado por Bamberger, Rugh e Mabry (2011), existe uma demanda crescente por avaliações sistemáticas de ações, programas e projetos governamentais.

Segundo os autores, cada vez mais acende o interesse de agências financiadoras, governos, sociedade civil, e até pelos próprios beneficiários, por respostas como: o projeto alcançou seus objetivos? O programa teve impacto sobre a população? Quem se beneficiou com a ocorrência do programa e quem não? O programa deve continuar? Respostas essas que só podem ser comprovadas por avaliações das políticas públicas.

Além da ampliação do interesse pelo campo da avaliação, o tema ainda encontra-se diretamente alinhado aos pressupostos mais atuais de modernização da gestão pública. Como afirmado por Dahler-Larsen (2005), o próprio movimento do *New Public Management* (NPM), ao demonstrar uma preocupação clara com a questão dos resultados (*outcomes*) das políticas públicas, define um importante papel para as avaliações das mesmas. O termo *avaliação* abrange um rico e dinâmico conjunto de abordagens focadas na efetividade, eficácia, eficiência de programas governamentais e, ainda, ressalta questões como a legitimidade de governos, a transparência e a *accountability*.

Para Dahler-Larsen (2005), o campo da avaliação de programas apresenta-se nos dias atuais como uma área de conhecimento relativamente autônoma, que conta com sua própria sociedade de profissionais, periódicos, conferências, cursos de treinamento, instituições e empresas de consultoria. O mesmo autor observa que o próprio setor público atenta para a importância de se construir uma estrutura organizacional para realizar avaliações e aparenta demandar, organizar e, ainda, influenciar ou controlar sua forma, método, duração e utilização. Cada vez mais é possível verificar a criação de institutos de avaliação relativamente independentes aos governos ou o estabelecimento de escritórios internos ao governo para a avaliação de determinados setores (DAHLER-LARSEN, 2005).

Prova do consenso sobre a importância das avaliações de programas e do vigor que o campo vem alcançando é a *onda* das atividades de avaliação experimentada por variados países em diferentes instâncias de governo. Consequentemente, a parcela do orçamento público destinada à contratação ou à confecção e acompanhamento de avaliações em governos tem aumentado de forma significante (DAHLER-LARSEN, 2005).

No Brasil, de forma semelhante, observa-se merecido crescente interesse nas últimas décadas devido, principalmente, à influência de organismos internacionais, à crise de financiamento do setor público e ao contexto de reforma do Estado, juntamente com a reorientação por uma

gestão pública por resultados (RIBEIRO, 2009). Calmon e Gusso (2002), ao apresentar a experiência recente do governo federal em avaliação, também compartilham do argumento dizendo que,

> [...] em meados da década de 1990, ganhou destaque o reconhecimento da necessidade de o governo [brasileiro] dispor de processos de monitoramento e avaliação que lhe permitissem controlar e articular suas ações e ganhar legitimidade por meio de eficaz comunicação de seus êxitos à sociedade (CALMON; GUSSO, 2002, p.7).

Dessa forma, o interesse pelas avaliações de programas no Brasil tem se ampliado nas últimas décadas, tanto no meio acadêmico, quanto na administração pública. É possível observar uma busca do aperfeiçoamento de programas com vistas a assegurar maior efetividade, bem como o fortalecimento de órgãos e unidades de auditoria, monitoramento e avaliação no setor público.

O mesmo movimento pode ser observado em governos estaduais, como é o caso do Governo do estado de Minas Gerais. Desde 2003, com a política do Choque de Gestão e a implantação de uma unidade de monitoramento intensivo dos programas estaduais e, a partir de 2007, com maior ênfase por meio do modelo Estado para Resultados, o governo mineiro introduziu um substantivo sistema de monitoramento e avaliação de suas políticas. Além disso, o governo ainda conta com o aparato da Fundação João Pinheiro (FJP), estrutura criada para subsidiar a administração pública estadual com avaliação de políticas públicas e produção de indicadores estatísticos, econômicos, demográficos e sociais dentre outras atividades.

Contudo, embora seja clara a percepção da importância das avaliações de políticas públicas, o crescimento do campo e o aumento dos investimentos realizados nessa área de avaliações, observa-se uma questão que merece destaque nesse debate: a utilização. Mesmo que exista um consenso inegável sobre o valor das avaliações de programas para a efetividade da administração pública em suas várias instâncias governamentais, pouca atenção tem sido dada à apropriação, pelos gestores dos programas, políticos e formuladores de políticas, dos resultados das avaliações. De acordo com Patton (2008), o modo como as avaliações são utilizadas afeta o gasto de bilhões de dólares investidos em programas para a redução da pobreza, melhoria das condições de saúde da população, políticas públicas de emprego e segurança pública, bem como outras áreas.

Embora a utilização de avaliações tenha sido um campo muito explorado internacionalmente, cujas pesquisas se iniciaram nos anos 1970 – quando os autores Guba e Lincoln (1989); Shulha e Cousins (1997); Weiss (1998); Patton (2008) passaram a questionar a efetiva utilidade das avaliações para influenciar decisões e contribuir para o aperfeiçoamento de programas e políticas – no Brasil, ainda não é perceptível tal preocupação. A ausência de estudos empíricos sobre a utilização das avaliações de programas, bem como a lacuna existente em modelos de meta-avaliações e carência por organizações que prezem por padrões nas avaliações são alguns fatos que provam a insuficiência brasileira nesse assunto (SERPA, 2010).

Nota-se que a maior parte dos estudos desenvolvidos no Brasil sobre avaliação se concentram no desenvolvimento de metodologias adequadas para avaliar políticas sociais, sendo escassos os autores brasileiros que ousaram tratar do tema da utilização dos resultados das avaliações. Ressalta-se que, mesmo que sejam desenvolvidos métodos sofisticados, de nada servirão se não forem utilizados para gerar modificações e aprimoramentos nos programas avaliados.

Nesse sentido, dada a insuficiência de literatura em língua portuguesa que trata do tema, atenta-se para a necessidade de formar um referencial coeso e completo sobre a utilização dos resultados das avaliações, seja para inserir o País no debate internacional, seja para ressaltar o desperdício de gastos com avaliação sem a apropriada utilização de seus resultados. Ademais, se no contexto do governo federal aponta-se carência para os estudos empíricos sobre utilização de avaliações, nos contextos estaduais eles inexistem. Não foi possível observar nenhum estudo, muito embora os entes da federação também tenham acompanhando o governo federal nas reformas na administração pública e tenham empreendido avaliações de suas próprias políticas.

Nesse contexto, apesar de ser clara a percepção de que a prática da avaliação de programas pode levar a uma transformação nas estruturas de conhecimento de organizações públicas, ou seja, na forma como ela detecta e corrige seus erros, pouca atenção tem sido dada à utilização das informações e resultados de programas governamentais. Apesar do crescente interesse em avaliações de programas governamentais no Brasil e nos governos estaduais, é possível observar certa escassez de estudos empíricos nacionais que permitam verificar a utilização de tais avaliações como instrumento de correção de rumos e aprimoramento dos serviços públicos.

Assim, torna-se necessário examinar, a partir da institucionalização da função avaliação no Brasil e em governos estaduais, como os sistemas avaliativos governamentais foram construídos, quais os objetivos das avaliações, suas características e, principalmente, se, de fato, favorecem a utilização de seus resultados. Tendo-se em vista tais considerações, apresenta-se o seguinte problema de pesquisa.

- *Como os diferentes stakeholders envolvidos nos processos de avaliação de políticas públicas realizadas pelo Governo do Estado de Minas Gerais percebem a utilidade dos resultados das avaliações? Por quê?*

1.2 Objetivos de pesquisa

Diante do problema apresentado, o objetivo geral desta pesquisa é *analisar a dinâmica da utilização dos resultados das avaliações de Projetos Estruturadores realizadas pelo governo do estado de Minas Gerais a partir do ano de 2007*. Em termos específicos, pretende-se:

1) identificar e analisar quais os fatores e mecanismos que influenciam a utilização (ou não) dos resultados das avaliações;

2) analisar o contexto e a conjuntura dos processos de implementação de avaliação de programas governamentais no Governo do Estado de Minas Gerais e quais são os principais resultados já identificados;

3) analisar como os instrumentos de avaliação têm contribuído para o aprimoramento e introdução de mudanças nas políticas avaliadas no contexto do governo do estado de Minas Gerais;

4) compreender como as abordagens adotadas nas avaliações de políticas públicas realizadas pelo governo do estado de Minas Gerais influenciaram a utilização de seus resultados;

5) propor alternativas de procedimentos para a utilização mais efetiva das avaliações dos programas governamentais.

Para fundamentar este projeto de pesquisa, será discutida, a seguir, parte do referencial teórico no tema da avaliação: histórico, conceito, finalidade e determinantes de seu uso em benefício das organizações públicas. A utilização insuficiente de avaliações, desintegrada ao ciclo de gestão, também será abordada, conforme será observado no capítulo seguinte.

2 AS AVALIAÇÕES DE PROGRAMAS GOVERNAMENTAIS

Patton (2008, p. 5) afirma que o tema avaliação emergiu da "[...] interface entre ciência e ação, entre o conhecimento e a atividade [e inaugurou questões fundamentais sobre] racionalidade humana, tomada de decisão e conhecimento aplicado à criação de um mundo melhor". Para o autor, o campo das avaliações não diz respeito somente aos pesquisadores da área, mas a um amplo público que se divide em políticos, administradores, beneficiários de programas governamentais, legisladores e cidadãos entre outros atores.

Este primeiro capítulo está dividido em três partes principais: primeiro, ao resgatar o histórico das avaliações de políticas públicas, expõe, de forma sucinta, o contexto em que o tema surgiu e se desenvolveu; na segunda seção, faz-se uma discussão sobre o conceito de avaliação na visão de diferentes autores e os tipos de avaliação mais utilizados, com um enfoque especial nas avaliações de impacto; por fim, na terceira e última seção do capítulo, pautado pela premissa de que nem todo estudo avaliatório será perfeito, mostram-se, de forma breve, as meta-avaliações, ou seja, as avaliações das avaliações, e alguns padrões da prática avaliatória, critérios criados para julgar sua a relevância e pertinência.

2.1 Avaliações: de onde vêm?

De acordo com Worthen, Sanders e Fitzpatrick (2004), o histórico da avaliação formal – o uso sistemático de informações e critérios para atribuir valores e justificar juízos – tem um caminho bem mais longo do que se pensa. Para os autores, a avaliação é evidente desde 2000 a.C. quando as autoridades chinesas faziam provas para os servidores públicos a fim de obter uma estimativa de seu desempenho ou quando Sócrates usava avaliações verbais para o processo de aprendizagem. No entanto, apesar de tais demonstrações, a história das avaliações formais no mundo ainda é um fato recente que teve seu crescimento associado ao *boom* das políticas sociais advindas dos regimes de *welfare state* de alguns países (WEISS, 1998).

A aplicação de métodos de avaliação de programas para as políticas sociais coincide com o crescimento e refinamento dos próprios métodos, bem com as mudanças ideológicas, políticas e demográficas que ocorreram no século XX (ROSSI; FREEMAN, 1998). Inicialmente, os métodos de avaliação foram aplicados às áreas de educação e saúde, principalmente, esforços relacionados às iniciativas de redução da mortalidade e de doenças infecciosas no período da

Primeira Grande Guerra Mundial. Nos anos de 1930, quando cientistas sociais de várias áreas defendiam o uso de pesquisas rigorosamente controladas nos programas sociais, as avaliações de políticas se tornaram mais frequentes. Nesse período, pesquisas na área social começaram a se tornar mais frequentes como, por exemplo, as experiências realizadas sobre a produtividade dos trabalhadores na experiência de *Hawthorne*, bem como estudos referentes ao *New Deal* americano.

De acordo Weiss (1998), o histórico da avaliação formal é um fato ainda recente, fortalecido na década de 1960, em decorrência da criação de programas para a resolução de problemas sociais, ao aumento do orçamento destinado a esses programas e, concomitantemente, à incapacidade de governos em auferir seus resultados e impactos. A partir daí, o tema avaliações tem observado considerável crescimento, acompanhando os debates relacionados com a rediscussão dos papéis do Estado.

Os programas sociais e as avaliações que os acompanham emergiram da transferência recente da responsabilidade das condições sociais e da qualidade de vida dos cidadãos, para os órgãos governamentais. Como Bremner (1956) descreveu, antes da Primeira Grande Guerra Mundial, exceto para os veteranos da guerra, a provisão de serviços sociais à população era vista como uma obrigação dos indivíduos e de associações voluntárias. Pessoas carentes, com problemas mentais e famílias desestabilizadas, eram vistas como clientes de instituições de caridade sustentadas por voluntários, geralmente mais afortunados.

Seguindo as consequências da Segunda Grande Guerra Mundial, numerosos programas na área social foram lançados com o objetivo de sanar os problemas de habitação, desenvolvimento urbano, educação e saúde advindos da guerra. Os gastos com os programas, cada vez maiores, acompanhavam as demandas das populações (ROSSI; FREEMAN, 1998). Também nesse período compromissos internacionais com programas de planejamento familiar, saúde, segurança alimentar e desenvolvimento rural foram firmados (ROSSI; FREEMAN, 1998).

Sem desconsiderar os esforços anteriores, foi durante os anos de 1960 que o número de estudos e artigos sobre avaliação de programas cresceu significativamente (WEISS, 1998, ROSSI; FREEMAN, 1998). Os cientistas sociais da época estavam engajados em analisar programas de prevenção de marginalidade, programas de habitação pública, atividades

educacionais, além de outras diferentes áreas. A monografia de Hayes (1959) sobre pesquisa em avaliação em países menos desenvolvidos e a revisão de Suchman (1967) sobre os métodos de avaliação são algumas ilustrações.

Como exemplo, pode-se citar a aprovação da Lei do Ensino Fundamental e Médio, em 1965, nos Estados Unidos, cujo propósito foi aumentar de forma considerável o financiamento desse país em educação (WORTHEN; SANDERS; FITZPATRICK, 2004). Contudo, mesmo com o crescimento considerável do orçamento educacional, não existia evidência convincente de que o aumento de qualquer financiamento no ensino tivesse gerado melhorias educacionais e, por isso, foram exigidos dos educadores que avaliassem suas próprias atividades por meio de um questionário que mostrasse a destinação dos recursos federais. De acordo com os atores, foi um começo marcado por dificuldades, em que milhares de educadores foram obrigados a analisar, sem qualquer experiência no assunto, dados empíricos quantitativos.

Como afirmado por Worthen, Sanders e Fitzpatrick (2004), ao final da década de 1960, o congresso americano aprovou a alocação de recursos para a avaliação de programas sociais outras áreas como, por exemplo, controle da criminalidade, tratamento do alcoolismo e saúde infantil, aumentando consideravelmente a demanda por avaliações.

No início de 1970, o campo da avaliação de programas emergiu como uma área especializada no interior das ciências sociais. Uma variedade de livros, incluindo as primeiras críticas quanto à qualidade dos estudos de avaliação (WEISS, 1972) e os problemas organizacionais associados à avaliação apareceram. Além disso, o avanço no campo de avaliação também pode ser atribuído à melhoria dos métodos de avaliação, à aplicação de estatísticas desenvolvidas e ao crescimento de programas eletrônicos e computadores.

De acordo com Rossi e Freeman (1998), a história da avaliação de programas pode ser, tanto obscura, quanto iluminada, dependendo da perspectiva adotada. Nos seus anos iniciais, as avaliações eram, muitas vezes, moldadas pelos interesses dos pesquisadores e cientistas sociais. Nos anos posteriores, no entanto, os clientes das avaliações tiveram significativa influência no campo. Atualmente, as avaliações são sustentadas por formuladores de políticas, planejadores dos programas e administradores que destinam parte de seu orçamento para as avaliações e acreditam no seu valor. Com o passar dos anos, esses vários clientes e *sponsors* das avaliações têm influenciado, cada vez mais, a definição do campo.

Apresentada, de forma breve, a história recentes das avaliações de políticas públicas no mundo, passa-se a embarcar no conceito de avaliação: o que são e qual a sua real finalidade na visão de diferentes autores, como se segue na seção seguinte.

2.2 Conceitos e abordagens

O tema avaliação acomoda muitas definições, com uma profusão de conceitos e significados (PATTON, 2008). Segundo Ala-Harja e Helgason (2000, p. 10) um significado comum a todas as avaliações seria: a constituição de "[...] análises sistemáticas e de aspectos importantes de um programa e de seu valor, de modo a fornecer conclusões confiáveis e utilizáveis".

Rossi e Freeman (1993) apontam que a avaliação é, antes de tudo, uma atividade complexa que envolve também decisões políticas e administrativas. Para os autores, avaliação é a aplicação de procedimentos sistemáticos de pesquisa social científica, com o objetivo de investigar a efetividade de programas de intervenção social para ajudá-los a aprimorar aspectos importantes como eficiência, desenho e implementação entre outros.

Para Cohen e Franco (1993, p. 77) o objetivo da avaliação é "maximizar a eficácia dos programas na obtenção de seus fins e a eficiência na alocação de recursos". Para Guba e Lincoln (1989), os objetivos de uma avaliação devem ir além dos objetivos da medida, da descrição e dos juízos, e adotar uma posição construtivista que toma em conta as interpretações e as significações que os atores elaboram da situação a avaliar. Alkin (2004) define avaliação como a atividade sistemática de obtenção de informações que podem ser utilizadas para mudanças de atitudes ou para a introdução de julgamentos que podem ser utilizados para o aprimoramento de projetos.

Tendo em vista esses conceitos, de acordo Weiss (1998, p. 4), cinco elementos podem ser percebidos em um conceito avaliação de programas: "[...] avaliação consiste na análise sistemática dos processos ou dos resultados de um programa (ou política), em comparação com um conjunto de padrões explícitos ou implícitos, com o objetivo de contribuir para o aprimoramento do mesmo".

O primeiro elemento (análise sistemática) enfatiza a metodologia da avaliação que, seja qualitativa, seja quantitativa, deve ser conduzida com rigor e formalismo. Na abordagem de Cohen e Franco (1993), a avaliação, apesar de similar, diferencia-se da pesquisa científica *pura* pela identificação de metodologias e técnicas próprias aplicadas e não somente ao incremento de conhecimento disponível.

Os segundo e terceiro elementos da definição (processos ou resultados) concentram-se no foco de investigação das avaliações que podem ser vários. Como relatam Cohen e Franco (1993), nem todas as avaliações são iguais. Podem-se estabelecer diferenças entre elas segundo vários critérios, tais como o tempo de suas realização, quem as realiza, a natureza que têm, a escala que assumem e a que alçada decisória são dirigidas.

O quarto elemento da definição de Weiss (1998) diz respeito ao julgamento, a comparação das evidências coletadas com o conjunto inicial de expectativas. Nesse sentido, a literatura sobre avaliações converge para o conceito que avaliar significa atribuir um valor, determinar a valia de algo. Para Arretche (1998), qualquer forma de avaliação envolve, necessariamente, uma forma de julgamento, uma medida de aprovação ou desaprovação.

Por fim, o quinto e último elemento aponta para a finalidade das avaliações de programas ou políticas: contribuir para o aprimoramento da política ou programa em questão. Nesse sentido, o propósito da avaliação não seria outro senão influenciar a reformulação de políticas antes de sua implementação (*ex-ante*), durante ou posteriormente (*ex-post*).

Com base no último elemento apresentado por Weiss (1998) e a partir da visão de vários outros autores (MOKATE, 2002; COSTA; CASTANHAR, 2003; PISCO, 2006), a avaliação é um instrumento de mudança que não deve apenas demonstrar os problemas ou propor soluções para os problemas detectados, mas produzir mudanças apropriadas que conduzam ao cumprimento de padrões de qualidade em diferentes serviços. Para Worthen, Sanders e Fitzoatrick (2004, p. 594), a avaliação, devidamente praticada, deve levar "[...] a melhorias diretas e incontestáveis de sistemas, programas e práticas – melhorias que não teriam ocorrido de nenhuma outra forma". Aguilar e Ander-Egg (1995, p. 31) sintetizam o conceito de avaliação, ao afirmarem que,

> A avaliação é uma forma de pesquisa social aplicada, sistemática, planejada e dirigida; destinada a identificar, obter e proporcionar de maneira válida e confiável dados e informações suficiente e relevante para apoiar um juízo sobre o mérito e o valor dos diferentes componentes de um programa (tanto na fase de diagnóstico, programação ou execução), ou de um conjunto de atividades específicas que se realizam, foram realizadas ou se realizarão, com o objetivo de produzir efeitos e resultados concretos; comprovando a extensão e o grau em que se deram essas conquistas, de forma tal que sirva de base ou guia para uma tomada de decisão racional e inteligente entre os cursos de ação, ou para solucionar problemas e promover o conhecimento e a compreensão dos fatores associados ao êxito ou ao fracasso dos resultados.

Para Ala-Harja e Helgason (2000), uma avaliação não se destina a abolir a subjetividade envolvida na tomada de decisões, mas permite certa melhoria no provimento de informações aos governantes, sobre as quais eles possam fundamentar suas decisões e melhor prestar contas sobre as políticas e os programas públicos. Como afirmam Cohen e Franco (1993, p. 278).

> [...] a avaliação não é um fim em si, e tampouco se pode pensar que seja o meio idôneo para superar todos os problemas que surjam na prestação de serviços [...]. Mas, sem dúvida ela contribui para aumentar a racionalidade na tomada de decisões, identificando problemas, selecionando alternativas de solução – dentro do possível – suas conseqüências, e otimizando a utilização dos recursos disponíveis.

Autores nacionais também reconhecem a importância da utilização das avaliações. Como assinalado por Calmon (1999), a avaliação é decisiva para o processo de aprendizagem organizacional. A conexão entre aprendizado organizacional e avaliação se estabelece a partir de um processo de *feedback*, no qual se percebem, monitoram e examinam informações relevantes advindas do ambiente interno e do externo. Patton (2008) argumenta que as avaliações devem ser focadas na utilização (*utilization-focused evaluation*) e disseminação dos resultados desde o início, do contrário, elas mesmas nem devem ser iniciadas.

Para Worthen, Sanders e Fitzpatrick (2004, p. 103), como qualquer disciplina jovem, "[...] a avaliação encontra controvérsias em torno de suas definições e ideologias". Autores mais ousados, como Henry (2000), no entanto, ultrapassam a questão da utilização e sugerem como objetivo final das avaliações *social betterment*, conceituado como a mudança social nas condições de vida da população. Para Henry (2000), que já havia postulado seu argumento em obras anteriores, a visão de que uma avaliação tem como objetivo a utilização de suas informações é reducionista e pode obscurecer seu verdadeiro propósito, qual seja, o de modificar a realidade social. O próprio autor, contudo, reconhece que alcançar o propósito de *social betterment* não é trivial. Indiretamente, as melhorias nas condições de vida da

população não podem ser evidenciadas sem antes os formuladores de políticas, políticos, técnicos de execução e os próprios cidadãos se apropriarem das informações geradas pelas avaliações e colocá-las em prática.

De acordo com Patton (2008), outra discussão que cerca o campo das avaliações diz respeito à diferenciação de avaliação para pesquisa científica, auditora e monitoramento. Para Cousins e Leithwood (1986), as avaliações nada mais são do que métodos de pesquisa social utilizados para responder questões colocadas por tomadores de decisão. De acordo com Weiss (1998), uma avaliação, de forma semelhante a qualquer pesquisa científica no campo social, irá aplicar métodos, tanto qualitativos, quanto quantitativos, tentará entender a relação entre variáveis e traçar uma relação causal entre as variáveis. Entretanto, para a autora, a diferença substancial entre ambas encontra-se na intenção, no objetivo pelo qual são realizadas: as avaliações destinam-se ao uso, e as pesquisas científicas colocam sua ênfase na produção de conhecimento. Patton (2008, p. 40) sintetiza bem essa ideia.

> A pesquisa científica básica é realizada para descobrir novos conhecimentos, testar teorias, estabelecer verdades, e generalizá-las no tempo e no espaço. A avaliação de programas é realizada para informar decisões, clarear opções, identificar fontes de aprimoramento, e prover informações sobre programas com limites de contexto e tempo, espaço, valores e política.

Dessa forma, as avaliações são desenhadas e conduzidas para um *cliente* que optou pela avaliação com o objetivo de gerar informação para uma tomada de decisões. Quanto ao monitoramento, Busjeet (2010) afirma que, embora a avaliação esteja fortemente ligada ao monitoramento, distingue-se dele por fazer juízos de valor sobre um programa com o uso de evidências, ao contrário do monitoramento, que, de forma periódica e contínua, utiliza as evidências para descobrir o progresso das ações de um programa em comparação com metas previamente acordadas.

De acordo com Patton (2008), as auditorias de programas e projetos no setor público diferenciam das avaliações, pois se destinam ao exame da conformidade legal dos mesmos. Como exemplo, o autor cita auditorias que têm por objetivo descobrir fraudes em campanhas políticas nos Estados Unidos. Assim, nas auditorias, o foco de fiscalização é mais presente que nas avaliações, uma vez que elas se destinam à análise do cumprimento de requisitos legais. A seguir, o quadro 1 sintetiza a diferença entre monitoramento, auditoria e avaliação de acordo com alguns critérios.

Quadro 1 – Diferenças entre monitoramento, auditoria e avaliação

Critérios	Monitoramento	Auditoria	Avaliação
Frequência	Periódico, regular. Contínuo, durante a implementação	Episódica, anual ou final	Episódica, a meio percurso, no final. Ocasional, durante e/ou após a implementação
Ação principal	Registro, supervisão, comparação com o plano	Conformidade às normas	Apreciação, medição dos efeitos, atribuição de causalidades
Objetivos	Modificar o plano de atividades ou de recursos. Intuito de melhorar a tomada de decisões sobre a gestão imediata das atividades que são acompanhadas	Garantir o respeito das normas, acabar com as práticas não conformes	Determinar a relevância, eficácia, eficiência, impacto e sustentabilidade de uma intervenção de desenvolvimento
Fontes de informação	Informações fornecidas regularmente pelo sistema de gestão	Informações de acompanhamento da gestão	Informações de acompanhamento, estudos, entrevistas, análises, inquéritos
Quem realiza	Chefe do projeto, responsável da política, setores de controle na organização	Auditores especialistas e acreditados (internos ou externos)	Avaliadores exteriores ao projeto ao à política (internos ou externos à organização) com a participação, em certos casos, da equipa do projeto

Fonte – Elaborado pela autora da dissertação.

Neste trabalho, utiliza-se o conceito de avaliação como o processo sistemático de coleta e análise de dados para determinar se os objetivos de um programa têm sido ou estão sendo alcançados e em que grau, com a finalidade de auxiliar na tomada de decisões. Nesse sentido, as avaliações existem para contribuir para o aperfeiçoamento de programas e políticas por intermédio da disponibilização de informações. A seguir, apresentam-se os tipos e abordagens de avaliações de políticas públicas mais comuns, embora seja difícil encontrar consenso nessa seara.

2.3.1 Tipos de avaliação

Muitas são as formas de diferenciar os tipos de avaliação. De acordo com Worthen, Sanders e Fitzpatrick (2004), diferenciadas concepções de modelos de avaliações foram criadas desde o surgimento dos primeiros estudos avaliatórios. De acordo com Aguilar e Ander-Egg (1995), mesmo que seja possível fazer diferentes classificações, recorrendo a diferentes critérios, quatro tipos de avaliações têm maior interesse prático: 1) segundo o momento em que se

avalia; 2) segundo o papel ou função da avaliação; 3) segundo a procedência dos avaliadores; 4) segundo os aspectos do programa que são objeto da avaliação.

De acordo com o primeiro critério apresentado por Aguilar e Ander-Egg (1995), segundo o momento em que se avalia, pode-se fazer uma tríplice distinção: avaliações *ex-ante*, avaliações intermediárias e avaliações *ex-post*. As avaliações *ex-ante*, também chamadas de iniciais ou pré-decisão, são realizadas entes de se realizar uma decisão de empreender um projeto ou programa, com a finalidade de proporcionar certa segurança e racionalidade para a decisão de implementar ou não um projeto. A avaliação intermediária, como o próprio nome indica, é feita durante a fase de execução de um projeto, fornecendo informações sobre o andamento do programa em relação ao planejamento inicial. Por fim, as avaliações *ex-post* são aquelas realizadas quando o programa chegou ao fim.

Quanto às funções que uma avaliação deve cumprir, têm-se as avaliações somativas e as avaliações formativas (AGUILAR; ANDER-EGG, 1995). Para Patton (2008), as avaliações somativas são conduzidas quando o programa já está implementado e são utilizadas para subsidiar decisões como, por exemplo, alocação de orçamento, continuidade de um projeto ou expansão entre outros. As avaliações formativas são orientadas ao aperfeiçoamento das políticas uma vez que são adotadas durante a implementação de um programa como meio de adquirir mais conhecimento sobre a gestão, desenvolvimento e resultado que se deseja melhorar.

O terceiro critério, procedência dos avaliadores, divide as avaliações em quatro tipos: externa, interna, mista e autoavaliação (AGUILAR; ANDER-EGG, 1995). Assim, a avaliação externa recorre a avaliadores que não pertencem à instituição responsável por executar o programa; a avaliação interna é feita com a participação de indivíduos (como avaliadores) que pertencem à instituição promotora do programa avaliado e a avaliação mista é uma combinação das anteriores. Na autoavaliação, as próprias pessoas envolvidas na execução do programa a ser avaliado participam das atividades avaliativas.

Existem vantagens e desvantagens relacionadas ao critério de procedência dos avaliadores de acordo com Aguilar e Ander-Egg (1995). Em uma avaliação interna ou na autoavaliação corre-se o risco de que os próprios avaliadores ressaltem aspectos positivos do programa em detrimento de aspectos que necessitem de melhoria, além da possibilidade de ocorrerem

conflitos pessoais entre os próprios integrantes da equipe na busca do responsável pelas falhas do projeto, caso sejam identificadas. Embora na avaliação interna exista maior segurança e familiaridade com o programa que está sendo avaliado, a garantia da objetividade é menor em consequência do envolvimento dos avaliadores com o objeto da avaliação. Na avaliação externa, por outro lado, os avaliadores apresentam maior imparcialidade quanto ao programa a ser avaliado, mas não detêm pleno conhecimento de todos os *fatores em jogo* de acordo com a lógica de funcionamento do programa e da organização em que o mesmo está inserido.

Por fim, para Aguilar e Ander-Egg (1995. p. 46), "[...] uma das classificações mais úteis é a que se faz atendendo aos aspectos ou componentes do programa que são objeto de avaliação", ou seja, quais aspectos dos programas devem ou não ser avaliados. As diferentes tipologias de avaliação que resultam desse critério têm implicações metodológicas e técnicas específicas para os avaliadores e demais envolvidos na avaliação.

De acordo com o critério *o que avaliar*, Figueiredo e Figueiredo (1986) apontam que a literatura de avaliação de políticas públicas costuma distingui-las em termos de sua efetividade, eficácia e eficiência. Por eficácia, a avaliação da relação entre os objetivos e instrumentos explícitos de um dado programa e seus resultados efetivos; por efetividade, entende-se o exame da relação entre a implementação de um determinado programa e seus impactos e, por avaliação de eficiência, a relação entre os custos empregados na implementação de uma dada política e os benefícios alcançados.

As avaliações de eficácia podem ser feitas entre, por exemplo, as metas propostas e as metas alcançadas pelo programa ou entre os instrumentos previstos para sua implementação e aqueles efetivamente empregados. Constitui, usualmente, o tipo mais aplicado nas avaliações correntes de políticas públicas, porque ela é a mais simples e menos custosa (FIGUEIREDO; FIGUEIREDO, 1986).

As avaliações de eficiência indicam uma noção de otimização dos recursos disponíveis, custo mínimo possível para o máximo de benefício (FIGUEIREDO; FIGUEIREDO, 1986). Para Arretche (1998), a avaliação de eficiência é a mais urgente de ser realizada, dado o contexto de escassez de recursos públicos que exige maior racionalização do gasto.

Por fim, as avaliações de efetividade (ou avaliações de impacto) constituem um tipo particular de avaliações que busca responder a perguntas de causa e efeito (KHANDKER; KOOLWAL; SAMAD, 2010). Ao contrário de avaliações gerais, que podem responder a perguntas a muitos tipos de perguntas, as avaliações de efetividade são estruturadas em torno do efeito causal (impacto) de um programa, isto é, as implicações do mesmo sobre determinado grupo de beneficiários. O foco na causalidade, marca das avaliações de impacto, determina as metodologias que podem ser utilizadas (KHANDKER; KOOLWAL; SAMAD, 2010). Para estimar o efeito causal de um programa sobre seu público-alvo, o método escolhido deve simular o que aconteceria com os beneficiários caso o programa não tivesse ocorrido, com o objetivo de descobrir quais mudanças nos beneficiários são diretamente atribuíveis ao programa em questão. Assim, para auferir o impacto de uma intervenção, utilizam-se diferentes métodos capazes de comparar os efeitos esperados do programa em um grupo de tratamento (grupos de indivíduos que participaram do programa) e outro grupo de controle (grupo de indivíduos com características semelhantes ao grupo tratamento, mas que não participaram do programa) em relação aos resultados/mudanças ao longo do tempo observados.

As avaliações de impacto também podem ser aplicadas a diferentes contextos, como afirmado por Gertler (2011) e abrangem problemas complexos, inerentes à natureza dos próprios programas sociais. Por exemplo, as avaliações podem ser direcionadas a programas de educação: qual o impacto das bolsas de estudo sobre o desempenho acadêmico? Ou programas de saúde: qual o impacto das campanhas de incentivo ao pré-natal sobre a redução da taxa de mortalidade infantil? Ou políticas de emprego: qual o impacto dos cursos de alfabetização de adultos na diminuição do emprego precário?

Assim, dada a natureza dos próprios programas avaliados e a complexidade das metodologias utilizadas nas avaliações de impacto, esse tipo de avaliação exige, normalmente, um tempo de execução mais prolongado e constituem avaliações mais custosas (ARRETCHE, 1998). Como afirmado por Garcia (2001), as avaliações de impacto enfrentam uma série de limitações, entre elas, a dificuldade de se isolar o impacto de uma política social, uma vez que se trata de contextos dinâmicos e interativos que se traduzem em distintos e inúmeros impactos.

No entanto, mesmo com a complexidade inerente às avaliações de efetividade, Aguilar e Ander-Egg (1995) argumentam que esse tipo particular de avaliações apresenta uma importância indiscutível.

> De que vale ou serve que a instituição responsável pelo programa funcione, se não se conseguem os resultados esperados? De que vale ou serve uma adequada implementação, se depois não se alcançam os resultados previstos? De que vale ou serve alcançar a cobertura prevista, se não se obtém o efeito esperado (AGULAR; ANDER-EGG, 1995, p. 55)?

Neste trabalho, o foco se dará em avaliações de impacto realizadas por avaliadores externos, dada a estratégia empreendida pelo governo de Minas Gerais a partir do ano de 2007, como será explicado no capítulo 4. A seguir, apresenta-se o conceito de meta-avaliação, ou seja, a avaliação da avaliação, com o objetivo de mostrar as limitações existentes nas avaliações e a necessidade, ilustrada por alguns autores, de criar critérios mínimos para diferenciar avaliações válidas, daquelas inúteis.

2.4 Como avaliar as avaliações?

Para Worthen, Sanders e Fitzoatrick (2004, p. 593) "[...] todo estudo avaliatório vai ser tendencioso em certa medida. As decisões que um avaliador toma sobre o que examinar – que métodos e instrumentos usar, com quem conversar e a quem ouvir – tudo isso influencia o resultado da avaliação." Para os autores, é importante que cliente e avaliadores tenham segurança e confiança das informações contidas nas avaliações: o cliente porque ele conta com informações relevantes e utilizáveis na avaliação contratada; e o avaliador, pois a sua reputação profissional estaria em jogo caso a avaliação fosse comprovadamente deficiente em algum ponto.

Tal como acontece em outros empreendimentos sociais, as avaliações podem ser boas, ruins ou algo entre esses dois extremos. De acordo com Stufflebeam (2001), dentre as muitas coisas que podem dar errado em avaliações, estão o foco inadequado, erros técnicos, custos excessivos, abuso de autoridade, implementação de má qualidade, relatórios atrasados, resultados ambíguos, interpretações inadequadas dos usuários, recomendações mal feitas entre outros. Por essas e outras razões, a meta-avaliação constitui, em essência, a avaliação da avaliação.

O termo meta-avaliação foi cunhado por Scriven (1969) e, a partir daí, tem absorvido diferentes definições e empregado de maneiras diferenciadas. Na visão de Worthen, Sanders e Fitzoatrick (2004, p. 594), a finalidade da meta-avaliação é "[...] ajudar a avaliação a realizar o seu potencial". Para os autores, se uma avaliação demonstra-se deficiente em algum ponto, o problema não se encontra no conceito da mesma, mas na forma como foi conduzida.

Para Stufflebeam (2001), a meta-avaliação apresenta o objetivo de julgar o mérito, a relevância, a qualidade, entre outros fatores, da avaliação por meio de um processo de questionamento sistemático. Segundo o autor, a tarefa de um meta-avaliador inclui tanto tarefas técnicas – como a análise do rigor metodológico e da confiança dos dados – quanto tarefas processuais e de interação com *stakeholders*. Assim, ao planejar uma meta-avaliação, o meta-avaliador deve se comunicar com os usuários da avaliação e outros públicos interessados para compreender e interpretar como a implementação e o processo avaliativo ocorreu. O meta-avaliador irá precisar fazer entrevistas ou *surveys*, além de outros tipos de coleta de informações e percepções dos indivíduos envolvidos e afetados pela avaliação original.

Para Worthen, Sanders e Fitzoatrick (2004), existem dois tipos de meta-avaliação: a meta-avaliação formativa tem como finalidade aprimorar o estudo antes que seja irremediavelmente tarde; a meta-avaliação somativa, por sua vez, objetiva aumentar a credibilidade dos resultados finais. Assim, como afirmado por Stufflebeam (2001), as meta-avaliações fomativas são empregadas antes mesmo do término das avaliações, com o objetivo de auxiliar os avaliadores a planejar, conduzir, interpretar e relatar os seus estudos avaliatórios de forma aprimorada; enquanto as somativas, conduzidas após a avaliação, têm a finalidade de ajudar a visualizar pontos fortes e as fraquezas da mesma, julgando seu mérito e valor.

Ao tratar do tema da meta-avaliação, pode-se nos referir indistintamente a outros conceitos similares, porém com aplicação diferenciada, como é o caso da meta-análise. A meta-análise é um tipo especial de análise que se destina a sintetizar resultados de múltiplos estudos (STUFFLEBEAM, 2001). Enquanto a meta-avaliação avalia a qualidade de um ou mais estudos de avaliação, a meta-análise tem por objetivo agregar e resumir os resultádos de vários estudos de avaliações qualitativas ou quantitativas. Para Weiss (1998), a meta-análise pode ser definida como um resumo sistemático dos resultados de uma série de avaliações de programas semelhantes para produzir uma estimativa do efeito global dos programas. Assim,

o objetivo primordial da meta-análise é fornecer uma estimativa generalizada do impacto de políticas que apresentam desenho e finalidade similares.

Neste trabalho, utiliza-se o conceito de meta-avaliação como a revisão sistemática de avaliações para determinar a qualidade de seus processos e resultados, de acordo com um conjunto de procedimentos, padrões e critérios para julgamento. No entanto, qual conjunto de padrões de conduta e qualidade que fornecem diretrizes para as avaliações deve ser considerado?

Preocupações com a ética, qualidade e utilidade das avaliações fundamentaram um esforço inicial pelos profissionais de avaliação para articular padrões da prática avaliatória (PATTON, 2008). Antes de 1970, de acordo com Patton (2008), os únicos critérios utilizados para julgar as avaliações eram aqueles baseados unicamente na metodologia e no rigor técnico, padrões idênticos aos utilizados para validar a qualidade do conteúdo de qualquer pesquisa científica no campo social. No entanto, no final da década, os próprios avaliadores perceberam que o rigor metodológico não sanava os problemas associados ao uso das avaliações.

De acordo com Stufflebeam (2001), atualmente, existem variados princípios e padrões de avaliação destinados a julgar o nível de aderência da avaliação. Como exemplo, podem ser citados o *Standards for educational and psychological testing* (1999), *Statistical Standards* (1992), *SEDCAR Standards* – para a condução de surveys de larga escala, GAO (1994) *Government auditing Standards*, o American Evaluation Association's Guiding Principles for Evaluators (1995) e o Joint Committee on Standards for Educattional Evaluation. Os padrões mais utilizados em meta-avaliações são os padrões de avaliação organizados pelo Comitê americano de padrões de avaliação educacional (Joint Committee on Standards for Educattional Evaluation) (JOINT COMMITTEE, 1994).

Os padrões do *Joint Committee* dividem-se em quatro grupos temáticos: 1) utilidade; 2) viabilidade; 3) adequação ética; 4) rigor ou exatidão, todos descritos resumidamente a seguir.

Os padrões de utilidade requerem que os avaliadores se familiarizem com as suas audiências, as definam claramente, verifiquem quais são, planejem-se para responder às carências informacionais dessas e relatem as informações relevantes em tempo hábil (JOINT COMMITTEE, 1994). Os padrões de utilidade guiam a avaliação para que ela seja

informativa, tempestiva e influente. Se seguidos esses parâmetros, a avaliação será conduzida para obter as informações essenciais e orientada pelo uso pretendido por aqueles que a tomarão para si.

Os padrões de viabilidade reconhecem que as avaliações definem e consomem recursos valiosos e que são geralmente executadas em um ambiente natural, não em laboratórios (JOINT COMMITTEE, 1994). A avaliação deve ter um desenho que seja operacional em campo, que não consuma mais fontes, materiais, pessoas ou tempo do que o necessário para responder às questões que a compõem. Os três primeiros padrões dessa categoria, tomados em conjunto, expressam se a avaliação é realista, prudente, diplomática e econômica.

Para facilitar a proteção dos direitos e do bem-estar dos indivíduos que se relacionam com a avaliação é que se estabeleceram padrões de adequação ética, que alertam contra a ilegalidade, a ação inescrupulosa, sem ética ou despropositada dos avaliadores (JOINT COMMITTEE, 1994). Essas diretrizes requerem que os avaliadores conheçam e sigam as normas referentes à privacidade, liberdade da informação e proteção dos seres humanos.

Por fim, os padrões de rigor ou exatidão determinam se uma avaliação produziu informações boas, relevantes e tecnicamente adequadas, bem como se os juízos estão explicitamente justificados pelos dados (JOINT COMMITTEE, 1994). A avaliação deve ser abrangente, isto é, considerar o máximo de características de identificação do programa e recolher dados sobre as características julgadas especificamente importantes para a avaliação do seu mérito. Os parâmetros de rigor ou exatidão objetivam garantir que a avaliação revele e transmita informações precisas sobre o mérito do programa, que produza e dissemine informações utilizáveis e válidas.

Quadro 2 – Padrões do Joint Committee on Standards for Educattional Evaluation (1994)

(continua)

Utilidade	
Identificação dos *stakeholders*	Pessoas envolvidas ou afetadas pela avaliação deveriam ser identificadas, para poderem ter suas necessidades respondidas
Credibilidade do avaliador	O condutor da avaliação deveria ser confiável e reunir as competências para realizar a avaliação, de modo que seus resultados alcancem a máxima credibilidade e aceitação
Alcance e Seleção da Informação	A informação coletada deveria ser selecionada para responder amplamente a questões pertinentes sobre o programa e ser responsiva às necessidades e interesses dos clientes e outros *stakeholders* envolvidos
Identificação dos valores	As perspectivas, os procedimentos e a racionalidade usada na interpretação das conclusões da avaliação deveriam ser cuidadosamente descritos, para estabelecer claramente as bases dos julgamentos de valor
Clareza dos relatórios	Os relatórios da avaliação deveriam descrever detalhadamente o programa, incluindo o contexto, os propósitos, os procedimentos e as conclusões da avaliação, de modo que a informação essencial seja provida e facilmente entendida.
Conclusão tempestiva dos relatórios e sua disseminação	Descobertas significativas no ínterim da avaliação e relatórios de avaliação deveriam ser disseminados para os usuários interessados, para poderem usá-los em tempo oportuno
Impacto da avaliação	Avaliações deveriam ser planejadas, conduzidas e reportadas de modo a encorajar os *stakeholders* a acompanhá-las e utilizá-las, de modo que a preferência pelo uso da avaliação aumente
Viabilidade	
Procedimentos pragmáticos	Os procedimentos da avaliação deveriam ser práticos, para minimizar as interrupções enquanto a informação necessária é obtida
Viabilidade política	A avaliação deveria ser planejada e conduzida com antecipação das posições dos diferentes grupos de interesse, para obter a cooperação destes, e identificar tentativas de qualquer um desses grupos de sabotar as operações da avaliação, ou enviesá-la, ou desorientar os resultados
Relações custo-benefício	A avaliação deveria ser eficiente e produzir informação de valor suficiente, de modo que os recursos despendidos possam ser justificados
Adequação Ética	
Orientação ao serviço	As avaliações deveriam ser desenhadas de modo a ajudar as organizações a endereçar e servir efetivamente às necessidades de todos os potenciais beneficiários do programa
Acordo formal	As obrigações formais das partes envolvidas na avaliação (o que fazer, como, por quem, quando) deveriam ser acordadas em contratos, de modo que essas partes estejam obrigadas a aderir a todas as condições do acordo ou formalmente renegociá-lo.
Direitos da pessoa humana	Avaliações deveriam ser desenhadas e conduzidas de modo a respeitar e proteger os direitos e o bem-estar dos indivíduos
Interação humana	Nas interações com outras pessoas, durante a avaliação, os avaliadores deveriam respeitar a dignidade humana e a autoestima dos demais, de modo que seus participantes não sejam ameaçados ou sofram danos
Avaliação completa e justa	A avaliação deveria ser completa e justa no exame e registro dos pontos fortes e fracos do programa avaliado, de modo que as forças possam ser aumentadas e as áreas problemáticas abordadas
Divulgação de resultados	Os formalmente envolvidos na avaliação deveriam garantir que todas as descobertas derivadas da avaliação se tornem acessíveis às pessoas afetadas pela avaliação e a quaisquer outros com expressos direitos legais de receber os resultados
Conflito de interesses	Conflitos de interesse deveriam ser negociados com abertura e honestidade, de modo que não comprometessem os processos da avaliação e seus resultados

Quadro 3 – Padrões do Joint Committee on Standards for Educattional Evaluation (1994)

(conclusão)

Responsabilidade fiscal	A alocação e o dispêndio dos recursos por parte do avaliador deveriam refletir procedimentos de responsabilização, ser prudentes e eticamente responsáveis, de modo que os gastos sejam contabilizados e apropriados
Rigor ou Exatidão	
Documentação do programa	O programa sob avaliação deveria ser descrito e documentado claramente e com precisão, de modo que o possa ser identificado nitidamente
Análise do contexto	O contexto no qual o programa se insere deveria ser examinado com aprofundamento suficiente, de modo que suas possíveis influências possam ser identificadas
Propósitos e procedimentos descritos	Os objetivos e procedimentos da avaliação deveriam ser monitorados e descritos em pormenor, de modo suficiente para que possam ser identificados e avaliados
Fontes de informação defensáveis	As fontes de informação utilizadas na avaliação do programa devem ser descritas em detalhe suficiente para que a adequação dessas informações possa ser avaliada
Informação válida	Os procedimentos para coleta de informações deveriam ser escolhidos (ou desenvolvidos) de modo que sua implementação assegurasse, para o uso pretendido, uma interpretação metodologicamente válida
Informação confiável	Os processos de obtenção de informação deveriam ser escolhidos ou desenvolvidos de tal modo que, quando implementados, pudessem assegurar que a informação obtida é suficientemente confiável para o uso pretendido
Informação sistematizada	A informação coletada, processada e reportada na avaliação deveria ser sistematicamente revisada e todos os erros encontrados deveriam ser corrigidos
Análise da informação quantitativa	Em uma avaliação, a informação quantitativa deveria ser apropriada e sistematicamente analisada, de tal modo que as questões da avaliação sejam efetivamente respondidas
Análise da informação qualitativa	Em uma avaliação, a informação qualitativa deveria ser apropriada e sistematicamente analisada, de tal modo que as questões da avaliação sejam efetivamente respondidas
Conclusões justificadas	As conclusões assumidas em uma avaliação deveriam ser explicitamente justificadas, para que os diversos grupos de interesse possam avaliá-las
Relatório imparcial	Os relatórios deveriam seguir procedimentos que os protegessem de distorções causadas por impressões pessoais ou parcialidades em qualquer parte da avaliação, de modo que os relatórios reflitam as descobertas com justiça
Meta-avaliação	A avaliação propriamente dita deveria ser formativa e resumidamente avaliada de acordo com esses ou outros padrões pertinentes, de modo que sua condução seja apropriadamente guiada e os interessados possam, ao fim, examinar suas forças e fraquezas

Fonte – JCSEE, 1994, adaptado pela autora da dissertação.

Assim, cabe afirmar que, como abordado nas seções anteriores, as avaliações têm como objetivo contribuir para aumentar a racionalidade na tomada de decisões, analisando problemas, definindo melhores intervenções de análise, otimizando a alocação de recursos. No entanto, Cohen e Franco (1999) argumentam que as avaliações não podem ser vistas como um fim em si e o único meio idôneo para superar todos os problemas envolvidos na elaboração e formulação de políticas públicas, uma vez que o tema está inserido em um contexto organizacional que, às vezes, pode não ser propício à avaliação. Weiss (1998)

menciona, como exemplo, que não faria sentido avaliar um projeto no qual os objetivos são ambíguos.

Embora não exista dúvida entre os atores do campo sobre a conveniência e importância em avaliar as políticas públicas, como declarado por Weiss (1998), cabe perguntar *para que* e *para quem* os projetos devem ser avaliados, pois, do contrário, podem acabar como estudos inúteis. Discutidos os conceitos de avaliação mais importantes e explicado o modelo desenvolvido pelo JCSEE, com seus quatro grupos temáticos e os parâmetros qualitativos que constituem cada um deles, passa-se para uma discussão mais profunda sobre a utilização das avaliações na visão dos autores que se dedicaram ao tema.

3 A UTILIZAÇÃO DOS RESULTADOS DAS AVALIAÇÕES

O debate sobre utilização das avaliações, inaugurado na década de 1970, expôs o *calcanhar de Aquiles* das avaliações de políticas públicas – as avaliações não estavam sendo utilizadas como deveriam (HENRY, 2000). Dahler-Larsen (2005) argumenta que o crescimento, a difusão e a institucionalização dos procedimentos de avaliação foram realizados antes mesmo que o problema mais fundamental sobre a utilização das avaliações fosse resolvido. Weiss (1972), ao listar uma série de casos em que as avaliações não estavam sendo utilizadas para a tomada de decisões, já declarava na época que "[...] a experiência no campo sugere que até os melhores estudos de avaliação de programas acabavam como lixo de uma usina burocrática" (WEISS, 1972, p. 11, tradução da autora da dissertação).

Este capítulo consiste em três partes: na primeira seção, discute-se o conceito de utilização; em seguida, apresentam-se as diferentes dimensões do uso das avaliações, ou seja, as formas como elas podem ser utilizadas; e, por fim, a última seção do capítulo trata de uma revisão das variáveis que influenciam (ou não) a utilização das avaliações na melhoria de programas públicos.

3.1 Discutindo a utilização

A questão da utilização dos resultados das avaliações tem sido discutida por uma série de autores a partir da década de 1970 (WEISS, 1972; PATTON, 1977; LEVITON; HUGHES, 1981; COUSINS; LEITHWOOD, 1986; ALKIN; COYLE, 1988; GUBA; LINCOLN, 1989). De acordo com Shulha e Cousins (1997), esses trabalhos foram responsáveis por inaugurar *o uso das avaliações* como um campo de pesquisa em potencial. A partir do esforço desses autores, uma série de outros trabalhos, datados da década de 1980, aprofundaram a pesquisa com apontamentos teóricos e empíricos.

De acordo com Patton (2008), não existe uma definição ideal e absoluta para o conceito de utilização, dado que a mesma é inerente aos valores e objetivos daqueles que farão uso dos resultados das avaliações. Por resultados das avaliações, entende-se qualquer informação associada aos achados de uma avaliação, sejam informações, interpretações, dados ou

recomendações comunicadas pelos avaliadores na conclusão de avaliações ou, até mesmo, no decorrer das mesmas (COUSINS; LEITHWOOD, 1986).

Para Alkin (2004, p. 19) *evaluation use* refere-se à proposta, às consequências planejadas que resultam da aplicação de informações de uma avaliação a um problema, questão ou preocupação. Em outras palavras, como demonstrado pelo autor, a utilização de avaliação alude à aplicação das informações geradas por uma avaliação para alcançar fins pretendidos. Para Shulla e Cousins (1997), a utilização das avaliações, ao contrário de um conceito unitário, representa um fenômeno multidimensional caracterizado por fatores como relevância da avaliação, credibilidade do avaliador, envolvimento com o usuário, comunicação, atendimento às necessidades de informação, percepção da avaliação como instrumento de gerenciamento e, principalmente, características do contexto da instituição. Assim, como afirmado pelos autores, a utilização pode ser vista como um fenômeno complexo aliado ao contexto, em que pessoas, ambiente e circunstâncias estão fortemente conectados, influenciando o uso.

Para Leviton e Hughes (1981), o termo utilização refere-se ao uso dos resultados das avaliações para os programas e políticas unicamente, excluindo a utilização por acadêmicos e pela imprensa, por exemplo. Para serem consideradas utilizadas, como afirmado pelos autores, é necessário, primeiro, que os resultados das avaliações possam ser traduzidos em recomendações para as políticas públicas avaliadas como refere Weiss *apud* Leviton e Hughes (1981) e, segundo, que existam evidências de que, na ausência da informação da avaliação, os envolvidos nas atividades do programa avaliado teriam pensado ou agido de forma diferente, o que se torna difícil de ser verificado por meio de pesquisas.

Segundo Leviton e Hughes (1981), é importante também diferenciar o conceito de impacto de uma avaliação do conceito de utilização. Por impacto entende-se a modificação nos programas e nas políticas advindas de contribuições das avaliações e, por outro lado, na utilização considera-se a relevância das descobertas da avaliação para questões que preocupam os envolvidos nos programas e políticas. Nesse sentido, as barreiras para o alcance dos impactos de uma avaliação estão relacionadas com três ocorrências: 1) avaliações podem ter baixa utilidade devido à baixa qualidade e praticidade das recomendações existentes; 2) a utilidade das avaliações podem não ser reconhecida ou os envolvidos podem escolher não utilizá-las; 3) uma avaliação pode não ter nenhum impacto dado que o mesmo é dependente

de decisões de diversos indivíduos governamentais que podem não enxergar utilidade nas recomendações das avaliações ou escolhem não utilizá-las.

Ultrapassando o conceito de utilização, autores como Alkin e Coyle (1988) iniciaram uma discussão com o objetivo de diferenciar os conceitos de não-uso (*non use*), má utilização (*misuse*) e abuso (*abuse*). Para os autores, existiria uma diferença muito grande entre a má utilização de uma avaliação, ou seja, utilização inadequada, e o não uso completo. Como exemplo, os relatório de avaliação não lidos e engavetados podem ser exemplos de não uso, mas não constituiriam exemplos de má utilização caso contenham informações incompletas e imprecisas impróprias ao melhoramento de um programa público.

De acordo com Patton (2008), os pesquisadores da utilização das avaliações enfrentam um desafio duplo: além de estudar e incentivar o uso adequado das avaliações, também trabalhar para eliminar os usos impróprios das mesmas. Mas o que constitui uma má utilização ou utilização inadequada ou imprópria?

Para Alkin e Coyle (1988), a definição de má utilização não constitui tarefa fácil. Por exemplo, ao considerar um caso de um administrador que esconde os resultados negativos de uma avaliação, para impedir que o público tenha acesso aos dados, pode-se presumir que se trata de um caso de má utilização (*misutilization*). No entanto, ao considerar que o mesmo administrador agiu dessa forma para não divulgar dados de uma avaliação inválida e mal-feita (*misevaluation*), pode-se considerar a atitude do administrador como responsável e apropriada, justificada como um não uso intencional.

Tanto administradores, quanto os próprios avaliadores estão relacionados com a questão da má utilização (ALKIN; COYLE, 1988). Por parte dos avaliadores, o uso incorreto de uma avaliação pode ser atribuído à má confecção do relatório, ao não atendimento das necessidades do usuário, bem como à falta de habilidade para a interpretação de dados quantitativos. Além disso, a escolha de metodologias e técnicas de coleta de dados inadequadas também constitui, na maioria dos casos, infração dos avaliadores. Quanto aos usuários, como se encontram envolvidos nas questões políticas das burocracias, espera-se que eles estejam mais direcionados a manipular os resultados de uma avaliação de forma intencional. Dessa maneira, enquanto à má utilização estaria mais atribuída aos usuários, uma avaliação inconsistente fica imputada à responsabilidade dos avaliadores.

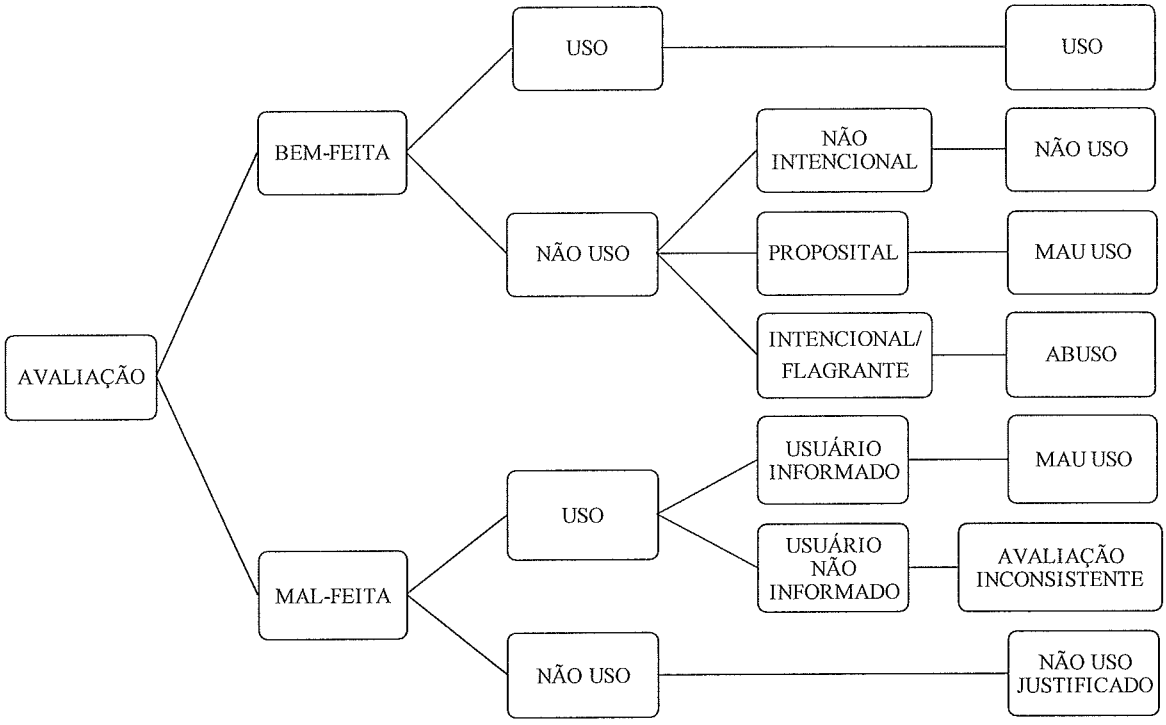
Outra diferenciação realizada por Alkin e Coyle (1988) diz respeito à má utilização intencional daquela não intencional. No primeiro caso, os usuários parecem estar cientes do uso incorreto. No entanto, quando, por exemplo, os usuários desviam informações dos relatórios de forma intencional para divulgação, tal atitude é entendida pelos autores como abuso da avaliação (*evaluation abuse*). No segundo caso, por outro lado, a utilização incorreta não intencional poderia ser atribuída à ignorância dos usuários, que não têm os conhecimentos metodológicos necessários. Neste caso, como os usuários não têm a obrigação de conhecer as metodologias de avaliação, os autores afirmam que se trata de uma avaliação inconsistente (*misevaluation*).

No diagrama proposto por Alkin e Coyle (1988), é possível verificar que a utilização das avaliações é um conceito que depende de fatores contextuais variados, conforme afirmado por Patton (2008). De acordo com os autores, a não utilização consciente nem sempre será considerada um problema, uma vez que tal ato pode ser justificado pela própria inconsistência da avaliação em questão (avaliações malfeitas) ou pela ausência de conhecimento necessário dos usuários.

Assim, como se pode observar na figura 1, existem dois tipos de avaliação relacionadas com o conceito de utilização de Alkin e Coyle (1988): as avaliações bem-feitas e aquelas malfeitas. Nas bem-feitas, o não uso pode ser: 1) não intencional, no caso, por exemplo, de desconhecimento dos usuários ou pouco aproveitamento dos relatórios; 2) proposital, quando as informações não são utilizadas, pois não condizem com os interesses dos usuários e demais *stakeholders* envolvidos no programa avaliado; 3) intencional ou flagrante, quando as informações das avaliações são corrompidas, perfazendo o caso de abuso das avaliações.

Nas avaliações malfeitas, o uso nunca será completo. Dessa maneira, mesmo que se faça um uso adequado das avaliações malfeitas, o caso sempre será considerado mau uso. Mesmo que o usuário seja informado, o será de maneira incompleta ou corrompida, dada a falta de qualidade da avaliação (figura 1).

Figura 1 – Relações entre a utilização das avaliações de Alkin e Coyle (1988)



Fonte – Alkin e Coyle, 1988, p. 335.

Além dos conceitos do não uso, Patton (2008) acrescenta a superutilização (*overuse*) que ocorre quando muita ênfase é destinada à utilização dos resultados de uma avaliação. Como exemplo, o autor cita a preocupação exagerada com os resultados de uma avaliação, geralmente fracas metodologicamente, tratando-os como se fossem definitivos ou impostos.

Neste trabalho, entende-se por utilização das avaliações o emprego dos resultados das mesmas – sejam informações, interpretações, dados ou recomendações, para a tomada de decisões governamentais. Como afirmado por diferentes autores (LEVITON; HUGHES, 1981, SHULLA; COUSINS, 1997, PATTON, 2008), acredita-se que a utilização é um tema complexo, relacionado a diversas variáveis que podem influenciar de forma positiva ou negativa o uso.

Discutidos os conceitos de utilização e não utilização, foca-se nas diferentes dimensões da utilização e como os autores foram, aos poucos, acrescentando novas abordagens ao longo do tempo. A utilização, como será visto a seguir, admite diferentes formas de uso, desde os mais tímidos, até aqueles que causam verdadeiras mudanças nos programas e pessoas envolvidos em uma avaliação.

3.2 As dimensões do uso

Quando os primeiros artigos sobre utilização das avaliações foram escritos no início da década de 1970, pensava-se, unicamente, que o uso das avaliações estaria relacionado com a tomada de decisões. Nesse sentido, como afirmado por Weiss (1972), entendia-se por utilização o uso dos resultados das avaliações para a tomada de decisões nos programas avaliados. Para a autora, como as avaliações são responsáveis por apontar recomendações sobre os processos ou os resultados de um programa, os avaliadores esperam que essas recomendações sejam aplicadas ao programa avaliado para modificá-lo, estendê-lo ou, até, finalizá-lo.

Contudo, conforme dito por Shulha e Cousins (1997), na década de 1980, já havia um consenso de que o uso, ao contrário de um conceito unitário, constitui um fenômeno multidimensional que pode ser mais bem descrito por diferentes dimensões descritas por Laura C. Leviton e Edward F. X. Hughes em 1981. Os autores distinguem três tipos de uso das avaliações, quais sejam, instrumental, conceitual e simbólico (ou político).

A dimensão instrumental diz respeito aos casos nos quais as conclusões das avaliações são utilizadas fundamentalmente para a tomada de decisões ou para a solução de problemas dos programas (LEVITON; HUGHES, 1981). Dessa forma, o conceito de utilização estabelecido por Carol H. Weiss em 1972 (utilização das avaliações para a tomada de decisões nos programas) constituía, assim, apenas uma das dimensões do uso, no caso, a dimensão do uso instrumental.

O uso instrumental é comum quando o avaliador entende bem o funcionamento do programa avaliado e suas questões, conduz a avaliação de forma adequada e comunica os resultados de forma satisfatória (WEISS, 1998). Para Weiss (1998), no uso instrumental, observam-se três condições: 1) as conclusões das avaliações não são controversas, nem provocam conflitos de interesse na estrutura da organização; 2) as mudanças a ser implementadas encontram-se dentro do escopo de atuação do programa e não provocam grandes rupturas no funcionamento do programa; 3) o ambiente do programa é relativamente estável, não apresentando dúvidas quanto ao seu orçamento, público-alvo e implementação entre outros. Uma quarta condição que favorece o uso instrumental trata da circunstância, do momento vivido pelo programa

quando ele recebe uma avaliação. Para Weiss (1998), quando o programa se encontra em uma situação de crise e ninguém sabe qual a melhor maneira para agir, a utilização instrumental de uma avaliação pode ser favorecida.

Uma segunda dimensão da utilização é o uso conceitual (LEVITON; HUGHES, 1981). O uso conceitual é circunscrito aos técnicos dos programas, cujo poder de decisão é limitado. Para Weiss (1998), mesmo que os técnicos sejam incapazes de utilizar os resultados das avaliações para a tomada de decisões no momento em que as conclusões das avaliações são apresentadas, as conclusões das avaliações podem modificar o entendimento desses técnicos sobre o que constitui o programa e como ele deveria funcionar. Assim, o uso conceitual, embora não se tenha uma decisão imediata decorrente de uma avaliação, possibilita a geração de novas ideias e direciona formas diferentes de ação para o programa. Dito de outra maneira, a dimensão conceitual caracteriza-se pela capacidade de influenciar a opinião dos técnicos a respeito de um determinado programa, sem que exista qualquer ação em decorrência dos resultados de uma avaliação. Contudo, como afirmado por Weiss (1998), quando as condições organizacionais se tornam favoráveis, os técnicos podem utilizar as novas ideias e conceitos apreendidos para introduzir mudanças nos programas de maneira instrumental.

Quando a avaliação é utilizada na dimensão simbólica ou política, os tomadores de decisões já têm conhecimento sobre o que seria necessário mudar em um determinado programa e, nesse sentido, as avaliações servem para justificar o processo de mudança (LEVITON; HUGHES, 1981). Nessa dimensão, os coordenadores de programas já detêm conhecimento das necessidades de seus programas. Reformulações e mudanças seriam implementadas de qualquer maneira, e as avaliações têm a função de legitimar o processo e ganhar adeptos, servindo como instrumento de persuasão (WEISS, 1998).

Além das três dimensões do uso citadas colocadas por Laura C. Leviton e Edward F. X. Hughes em 1981 (instrumental, conceitual e simbólica ou política) um debate extenso aponta a existência de mais dimensões. Patton (2008) relata que, na medida em que o campo da avaliação foi sendo ampliado e novas discussões sobre utilização surgiram, distinções e tipologias adicionais emergiram.

Weiss (1998), por exemplo, aponta uma quarta dimensão do uso das avaliações conceituada como *enlightenment* e traduzida para o português como *esclarecimento*. Para a autora, as

avaliações podem ser utilizadas para influenciar outras instituições ou eventos que vão além do escopo de determinado programa avaliado. A quarta dimensão se justifica uma vez que as conclusões de uma avaliação podem ultrapassar seus limites originais e pautar, não somente os coordenadores de programas avaliados, mas também uma determinada comunidade acadêmica e coalizões políticas interessadas nos resultados de dada política pública. igualmente é possível que os resultados das avaliações mudem a maneira de pensar determinada política pública ou política social, alterando paradigmas sobre os seus benefícios (WEISS, 1980). Nesse caso, a avaliação acarreta impacto sobre os formadores de opinião, bem como alterações nas crenças e na forma de ação de instituições modificando, assim, conhecimentos sobre a formulação de políticas e agenda governamental.

A quarta dimensão da avaliação, contudo, não ocorre de forma banal (WEISS, 1998). Entretanto, como argumentado pela autora, a acumulação de conhecimento é uma importante força na formulação de políticas. Quando uma avaliação contribui para o acúmulo de conhecimento, pode colaborar para escalas de modificação de pensamento e, em última instância, em mudanças nos cursos de ação das políticas. Nem todos os avaliadores aspiram esse tipo de influência para os resultados das avaliações, uma vez que representa uma modificação maior que a instrumental. No entanto, de acordo com Weiss (1998), alguns casos demonstram que *enlightenment* não é uma categoria desprezível.

Patton (2008) adiciona mais uma forma de utilização, cunhada pelo autor, de uso processual. Para o autor, a utilização pelo processo ocorre quando pessoas envolvidas nas avaliações (com exceção dos avaliadores) demonstram alterações comportamentais e cognitivas, como resultado da participação no processo de avaliação. Essas mudanças são reflexo, portanto, da aprendizagem experimental adquirida ao longo da realização de uma avaliação e da convivência com avaliadores, apreendendo seus valores e forma de raciocínio. Em síntese, o uso processual envolve "[...] aprender a pensar como avaliador" (JOHNSON, 1998, p. 93) e pode ter retornos positivos no longo prazo como, por exemplo, via habilidades aprimoradas, entendimento da linguagem de avaliação, maior aproveitamento dos resultados das avaliações, melhor tomada de decisões e propriedade e segurança no acompanhamento de novas avaliações. O uso pelo processo, dessa forma, se sobrepõe ao uso instrumental e conceitual.

Além do uso processual, Patton (2008) adiciona mais quatro tipologias de utilização: legitimativo, persuasivo, imposto e mecânico. O uso legitimativo ocorre quando os resultados de a avaliação são usados para dar suporte a uma decisão e justificá-la, decisão esta que seria implementada antes mesmo da avaliação ser realizada. O uso persuasivo refere-se à utilização seletiva dos resultados da avaliação para dar suporte às decisões específicas sobre financiamento e debates políticos. O uso imposto acontece quando indivíduos do alto escalão, com maior nível de autoridade, obrigam a utilização àqueles situados em um nível hierárquico menor. Por fim, o uso mecânico, ou uso por conformidade, refere-se ao simples cumprimento das exigências de uma avaliação, uma vez que a implementação dos resultados de uma avaliação ocorre de forma forçosa, condicionada, por exemplo, a aumentos de orçamento para o programa que o fizer.

Assim, Patton (2008), ao reunir as tipologias clássicas às mais atuais, cria uma classificação para as possíveis dimensões do uso, quais sejam: 1) uso direto – instrumental, conceitual e processual; 2) uso no longo prazo – *enlightenment* e influência; 3) uso político – simbólico, legitimativo, persuasivo, imposto e mecânico, como demonstrado no quadro 3.

Quadro 4 – Diferentes usos das avaliações

(continua)

Categoria (Patton)	**Tipos de Uso**	**Descrição**
Uso direto	Instrumental	Quando os tomadores de decisão utilizam os resultados da avaliação para modificar o objeto da avaliação, de alguma forma
	Conceitual	A avaliação possui capacidade de influenciar a opinião dos técnicos a respeito de um determinado programa, sem que exista qualquer ação em decorrência dos resultados de uma avaliação
	Processual	A avaliação tem a capacidade de impor mudanças advindas do engajamento no processo de avaliação e no aprendizado de *pensar como avaliador*
Uso no longo prazo	*Enlightenment*	A avaliação é utilizada para influenciar outras instituições ou eventos que vão além do escopo de determinado programa avaliado
	Influência	A avaliação possui *unaware/unintended* impactos, ou seja, impactos não previstos ou planejados antecipadamente pelos avaliadores
Uso político	Simbólico	A avaliação é utilizada na dimensão simbólica ou política os tomadores de decisão já têm conhecimento sobre o que seria necessário mudar em um determinado programa e, nesse sentido, as avaliações servem para justificar o processo de mudança
	Legitimativo	Os resultados de uma avaliação são utilizados como forma de justificar uma decisão que seria tomada
	Persuasivo	Utilização seletiva dos resultados da avaliação para dar suporte a decisões específicas sobre financiamento e debates políticos
	Imposto	Indivíduos do alto escalão, com maior nível de autoridade, obrigam a utilização das avaliações àqueles situados em um nível hierárquico menor

Quadro 5 – Diferentes usos das avaliações

(conclusáo)

Uso político	Mecâncico	Refere-se ao simples cumprimento das exigências de uma avaliação, vez que a implementação dos resultados de uma avaliação ocorre de forma forçosa, condicionada, por exemplo, a aumentos de orçamento para o programa que o fizer. Geralmente a implementação das próprias avaliações ocorre por obrigação e não por interesse dos usuários

Fonte – Patton, 2008, p. 112, adaptado pela autora da dissertação.

Torna-se importante ressaltar que uma dimensão raramente é encontrada sozinha, como afirmado por Weiss (1998). Muitas vezes, é dificultoso identificar a dimensão de uso em que uma avaliação é encontrada, já que os diferentes tipos de uso podem ser permutados em um mesmo estudo avaliatório.

Descritos os conceitos de utilização de avaliação e mencionadas as diferentes dimensões do uso, a seção seguinte apresenta os fatores que influenciam a utilização das avaliações nos programas e nas instituições. Nesse sentido, como se segue, serão apresentados critérios e circunstâncias que afetam a utilização das avaliações na visão de diferentes autores.

3.3 As vaiáveis que influenciam a utilização das avaliações

Na revisão feita por Shulha e Cousins (1997), os autores apontam que, em 1986, já havia um consenso sobre a utilização inadequada das avaliações e sobre a necessidade de ultrapassar a simples catalogação dos diferentes usos de avaliação (instrumental, conceitual e simbólico). Nesse período, buscavam-se modelos para a mensuração dos diferentes fatores e variáveis que influenciavam a utilização das avaliações como, por exemplo, a relevância do contexto em que a pesquisa avaliatória está inserida. Como consequência, inúmeros estudos começaram a desenvolver teorias e desenhar estratégias na tentativa de sanar a não utilização, a má utilização e o abuso das avaliações.

Neste trabalho, destacam-se três estudos principais que contribuíram para a pesquisa sobre as variáveis que influenciam a utilização das avaliações: 1) a pesquisa de Cousins e Leithwood (1986), que tiveram por objetivo identificar os estudos empíricos sobre uso das avaliações conduzidos entre 1971 e 1985, destacando-se o trabalho de Leviton e Hughes (1981); 2) o *paper* de Shulha e Cousins (1997), que consolida uma análise semelhante à já realizada em

um período diferente, dessa vez de 1986 à 1997; e, por fim, 3) o estudo de um grupo de pesquisadores (JOHNSON *et al.*, 2009) que se dedicou a uma pesquisa sobre a utilização das avaliações nos principais periódicos de avaliação em um período de 25 anos compreendido entre 1986 a 2005.

3.3.1 Cousins e Leithwood (1986): a importância do contexto

Para Cousins e Leithwood (1986), o conhecimento advindo da experiência prática fornece a base para a ação e decisão na maioria das organizações. O conhecimento derivado de estudos e métodos aplicados às ciências sociais, por outro lado, tende a ser seletivo, pouco compartilhado e, muitas vezes, apenas serve de complemento ao conhecimento prático. Motivados por esse pensamento, os autores buscaram revisar as variáveis (dependentes e independentes) que influenciavam a utilização de informações de uma avaliação.

Cousins e Leithwood (1986) identificaram 65 estudos empíricos sobre uso das avaliações conduzidos entre 1971 e 1985, por meio da busca por estudos que continham os termos *evaluation utilization*, *data use*, *desision making* e *knowledge utilization*. Quatro perguntas principais guiaram a pesquisa dos autores: 1) quais as características metodológicas dos estudos empíricos para investigar a utilização das avaliações e os seus determinantes? 2) como as variáveis dependentes foram operacionalizadas nos 65 estudos empíricos encontrados? 3) quais as orientações em relação às variáveis independentes foram adotadas? 4) que fatores têm demonstrado influência na utilização dos resultados da avaliação?

Como definição de utilização das avaliações, Cousins e Leithwood (1986) consideraram duas variáveis dependentes: uso de avaliações como apoio para decisões e uso educativo das avaliações para o aprendizado dos tomadores de decisões. Como variáveis independentes, por sua vez, os autores consideraram os fatores que influenciam os tomadores de decisões a utilizar as avaliações. Dos 65 estudos analisados, Cousins e Leithwood (1989) observaram que, independentemente do tipo de uso da avaliação (pata tomada de decisões ou para aprendizado), diversas foram as orientações coerentes com a visão de que o uso corrente das avaliações era limitado.

O exame realizado pelos autores resultou em duas categorias de fatores relacionados à utilização das avaliações: características de implementação e características da decisão ou do

policy setting. As características de implementação foram desdobradas em seis: qualidade da avaliação, credibilidade, relevância, qualidade da comunicação, resultados e tempestividade. As características da decisão ou do *policy setting*, por sua vez, também foram desdobradas em seis: necessidade da informação, características da decisão, clima político, informação competitiva, características pessoais e comprometimento e receptividade à avaliação.

Quadro 6 – Fatores relacionados à utilização das avaliações na pesquisa de Cousins e Leithwood (1986)

Categoria	Características
Características de implementação	Qualidade da avaliação
	Credibilidade
	Relevância
	Qualidade da comunicação
	Resultados
	Tempestividade
Características da decisão ou do *policy setting*	Necessidade da informação
	Características da decisão
	Clima político
	Informação competitiva
	Características pessoais
	Comprometimento e receptividade à avaliação

Fonte – Cousins e Leithwood, 1986, adaptado pela autora da dissertação.

De acordo com Cousins e Leithwood (1986), mais de 40% dos estudos analisados por eles observaram, em maior ou menor grau, relações entre o uso das avaliações e a sua qualidade. Por qualidade das avaliações entendem-se a sofisticação metodológica, o tipo de abordagem utilizada ao problema da avaliação, ou a intensidade das atividades avaliativas. No entanto, as pesquisas do período não chegam a um consenso, enquanto alguns atores (YEH, 1980) apontavam que métodos menos sofisticados proporcionam um uso melhor das avaliações, outras linhas indicavam que a robustez metodológica tem uma relação positiva com o uso (PATTON et al., 1977). Outros autores, como Leviton e Boruch (1983), argumentaram que estudos avaliatórios focados em desenho ou processos seriam mais bem utilizados que estudos focados em resultados dos programas. Por fim, como intensidade das atividades avaliativas, os autores mostram que os estudos avaliatórios que fizeram uma pesquisa mais ampla, com maior número de testes e/ou observações, apresentaram melhor utilização.

Quanto à variável credibilidade, os autores observaram que 25% dos estudos analisados citavam essa variável como influente na utilização (COUSINS; LEITHWOOD, 1986). Quando a avaliação era entendida pelos tomadores de decisões como m instrumento válido e

de singular importância para as atividades desempenhadas pela organização, o uso, ou o potencial para a utilização parecia maior.

Os 18 estudos que exploraram a relação entre relevância e a utilização das avaliações conceituaram a variável como uma medida do direcionamento da avaliação aos usuários, ou seja, o grau de importância da avaliação para os públicos que a receberam (COUSINS; LEITHWOOD, 1986). Os estudos analisados demonstraram que aquelas avaliações que refletem o conhecimento do contexto em que as mesmas estavam inseridas seriam mais bem utilizadas que aquelas afastadas do contexto da organização em que se inserem. Como uma medida da relevância de uma avaliação, a posição organizacional do avaliar (interna ou externa), tem direta relação com a utilização. A presença de avaliadores internos sugeriria uma maior utilização por eles apreenderem maior conhecimento sobre as características dos programas avaliados e a cultura da organização em que estão inseridos.

Embora um terço dos estudos indicasse que a qualidade da comunicação na apresentação dos resultados das avaliações e a amplitude da divulgação mostram alguma relação com o uso, os mesmos não se mostraram conclusivos (COUSINS; LEITHWOOD, 1986). Observou-se que, ainda que houvesse apresentações orais dos resultados e recomendações das avaliações (juntamente com os relatórios), com uma linguagem *não-técnica*; esse fato, no entanto, pouco contribui para a tomada de decisões em um programa. Em contraste, a utilização pareceu estar fortemente relacionada à comunicação permanente e à proximidade (geográfica) entre avaliador e tomador de decisões.

Quanto à variável resultados, 19 estudos forneceram evidências sobre os efeitos dos resultados no uso das avaliações (COUSINS; LEITHWOOD, 1986). A maioria desses estudos relatou que descobertas esperadas estariam mais relacionadas à aceitação e à utilização por partes das equipes dos programas, enquanto resultados surpreendentes das avaliações estariam fadados ao não uso. Recomendações mais práticas e conclusivas também mostraram-se mais receptivas ao uso que resultados gerais e pouco diretos.

Por fim, como última variável relacionada às características de implementação, a questão tempo, ou seja, o fornecimento oportuno e tempestivo dos resultados das avaliações foi positivamente relacionado à utilização em menos de 15% dos estudos analisados (COUSINS;

LEITHWOOD, 1986). Patton et al. (1977), no entanto, mencionaram que o atraso em prazos e na entrega dos relatórios de avaliações pouco influencia a sua utilização.

Na parte direcionada às características da decisão ou do *policy setting*, as avariáveis categorizadas de acordo com as necessidades de informação foram variadas (COUSINS; LEITHWOOD, 1986). Elas incluíram, tanto a intensidade da necessidade, que o tipo de informação necessária e variâncias em relação ao público solicitador das informações. Quanto à intensidade da necessidade, observou-se que, quanto maior precisão, melhor a utilização. Quanto ao tipo de necessidade, verificou-se que necessidades de informação voltada para a ampliação de programas apresentam um uso mais positivo das informações. E, quanto ao público, maior variedade de audiências quanto aos resultados das avaliações mostrou-se influenciar de forma negativa os resultados das avaliações.

Mais de um terço dos estudos relatou relação positiva entre a utilização e as características da decisão como, por exemplo, área, contexto e significado da decisão. Cousins e Leithwood (1986) relataram que a maior parte dos trabalhos analisou a utilização para diferentes áreas da decisão, sendo que, nas áreas em que as avaliações apontavam necessidades como capacitação, pontos fracos do programas e questões ligadas à administração do programa, o uso foi caracterizado como elevado. Contudo, em contraste, decisões mais significativas, isto é, que envolvem mais atores e são mais conflituosas, mostraram relações diversas com a utilidade das avaliações.

A importância do clima político foi considerada tanto em nível organizacional, quanto extraorganizacional, sendo observado apenas 10 estudos (COUSINS; LEITHWOOD, 1986). De acordo com os autores, a utilização diminui caso seus resultados sejam percebidos como uma ameaça pelos funcionários da organização. Como exemplo de influência extraorganizacional, os autores citam o exemplo de exigências de realização de avaliações para o financiamento de projetos como um dos fatores que influenciam positivamente o uso.

Quando confrontados com um problema, os tomadores de decisões confiam em várias fontes de informações diferenciadas, além dos resultados das avaliações. Por isso, Cousins e Leithwood (1986) consideraram a variável informação competitiva, indicando 14 estudos que constataram fontes alternativas de informação competindo com os resultados da avaliação para tomada de decisões. O conhecimento prático de pessoas envolvidas nos programas,

crenças, valores e interesses, por exemplo, mostraram-se concorrentes poderosos com os resultados das avaliações.

Mais de um terço dos estudos analisados por Cousisn e Leithwood (1986) relataram relações entre as características pessoais do tomador de decisões, como liderança, experiência, treinamento e papel na organização e utilização das avaliações. No entanto, de acordo com os autores, as informações são conflitantes. Poucos estudos apoiaram a relação entre posição do tomador de decisões na hierarquia da organização e a adequada utilização das avaliações, ou seja, a ideia de que uma posição mais alta na hierarquia seria favorável ao uso. A relação entre fatores pessoais, como experiência, também mostrou-se ambígua, a experiência no campo contribui mais para diferentes formas de processamento das informações. Características de liderança mostraram-se positivamente associadas ao uso das avaliações.

O comprometimento e receptividade à informação da avaliação foi definida em termos do grau em que o tomador de decisões estava envolvido na avaliação, sendo que, 25% dos estudos analisados afirmaram que existe uma relação direta entre a participação e acompanhamento do processo avaliatório e a adequada utilização das avaliações (COUSINS; LEITHWOOD, 1986).

Por fim, descritas todas as variáveis que influenciam a utilização das avaliações, Cousins e Leithwood (1986) sintetizam os fatores que, no geral, parecem causar mais impactos sobre o uso das avaliações. Assim, a utilização das avaliações parecia ser mais fortemente evidente quando:

- os tipos e abordagens avaliatórias foram adequadamente selecionados, juntamente com a metodologia e a intensidade em que a avaliação ocorreu;
- as decisões a serem tomadas tinham relevância para os usuários e necessitavam de coleta de dados analisados por avaliações;
- os resultados das avaliações foram condizentes com as crenças e expectativas dos usuários;
- os usuários foram envolvidos no processo de avaliação e tinham um compromisso com os resultados anteriormente ao processo avaliatório;
- os usuários consideraram as informações relatadas por as avaliações serem relevantes para suas preocupações e problemas;

- não existirem, ou foram mínimas, as fontes de informações concorrentes com as avaliações que colocaram em conflito os resultados.

No período estudado pelos atores (1971 a 1985), um estudo em particular merece destaque. Trata-se do primeiro estudo que, de fato, objetivava mensurar o grau de utilização das avaliações, como se observa a seguir.

3.3.1.1 Leviton e Hughes (1981): o primeiro passo em busca das variáveis de utilização

Apesar das várias discussões sobre os diferentes tipos de utilização das avaliações, Leviton e Hughes (1981) foram os primeiros a se preocupar com a mensuração do grau de utilização das mesmas e com as variáveis que influenciam o processo. No artigo escrito em 1981, cujo objetivo foi revisar as definições para o estudo da utilização das avaliações, os autores discutem criticamente o conceito de utilização e a metodologia utilizada para auferir a utilização e, por fim, fazem uma revisão das variáveis que afetam a utilização.

Na parte destinada às considerações metodológicas, Leviton e Hughes (1981) argumentam que, tanto nos estudos caso, quanto nas entrevistas ou *surveys*, o estudo sobre a utilização de avaliações padece, no mínimo, de quatro grandes problemas. Primeiro, é sempre difícil documentar quando e como a utilização ocorre, dado que as avaliações são frequentemente utilizadas de maneira informal e, também, porque pode haver um lapso entre a utilização em si e o estudo da avaliação. Além disso, estudos retrospectivos da utilização de avaliações podem focar poucos e dramáticos usos ao contrário daqueles modestos e mais frequentes. Um segundo problema refere-se à dificuldade de demonstrar que a mudança, em qualquer nível de organização, foi causada, ao menos em parte, por avaliações. Um terceiro problema está relacionado à ausência de medidas utilizadas para apurar a taxa de utilização. E, por fim, o quarto problema envolve a unidade de análise: qual seria um exemplo adequado de utilização? Vários tipos de utilização podem estar presentes na produção de impacto de uma avaliação.

Como conclusão, Leviton e Hughes (1981) encontraram cinco variáveis principais para auferir a utilização das avaliações: 1) a relevância da avaliação para as necessidades dos usuários em potencial; 2) o grau de comunicação entre avaliadores e usuários; 3) tradução de avaliações e suas implicações para os programas; 4) credibilidade ou confiança depositada nas avaliações; 5) o compromisso de usuários individuais.

A primeira variável – relevância da avaliação – estaria relacionada à necessidade dos clientes por informações e ao conteúdo das avaliações. De acordo com Leviton e Hughes (1981), estudos anteriores demonstram que baixa relevância das informações das avaliações comprometia a utilização. No entanto, duas são as dificuldades relacionadas à identificação de estudos avaliatórios relevantes: atender as necessidades dos clientes de uma avaliação, sejam eles gestores ou tomadores de decisões (políticos), e a tempestividade da informação gerada. Quanto aos clientes, os autores afirmam que, enquanto para os políticos as considerações mais importantes das avaliações estão relacionadas à efetividade geral do programa, para os gestores são as informações sobre a efetividade de diferentes elementos do programa as mais relevantes. Quanto à tempestividade, enfatiza-se a importância de submeter os resultados de uma avaliação em tempo de se tomar as decisões necessárias, principalmente para o uso instrumental. Como exemplo, os autores citam o valor de entregar as avaliações ao congresso durante as fases corretas do processo legislativo.

Como segunda variável, Leviton e Hughes (1981) determinaram que um estreito processo de comunicação entre os consumidores e os avaliadores distingue estudos avaliatórios bem-sucedidos daqueles que não foram utilizados. No entanto, como afirmado pelos autores, a comunicação tende a ser obstruída pela burocracia excessiva. Como exemplo, os autores citam a hierarquia presente nas organizações públicas que pode impedir que as informações cheguem aos potenciais usuários.

Além disso, as informações das avaliações podem ser distorcidas ou até eliminadas durante o processo de comunicação dos resultados. Nesse caso, a utilidade da avaliação pode ser enviesada, já que os consumidores da avaliação podem ter uma ideia alterada ou incompleta, que podem modificar todo o foco da utilização dos dados de uma avaliação. Isso pode acontecer, por exemplo, na transformação dos relatórios completos de avaliação para os resumos executivos dos mesmos que serão lidos, geralmente, pelos administradores do alto escalão.

Ainda, quanto ao quesito comunicação, os autores também observam a importância da disseminação de informações, ou seja, o formato da apresentação dos resultados das avaliações. Como notado por Weiss (1978), enquanto a disseminação formal de informações é

rara nas burocracias, a condução informal de informação entre as agências é frequente como parte de um *colégio invisível* de especialistas.

A variável seguinte à comunicação se trata do processamento de informações, conceito por Leviton e Hughes (1981) como o reconhecimento dos resultados da avaliação e a sua tradução em recomendações e ações implementáveis. Para os autores, a maneira como os resultados das avaliações são apresentados aos usuários afeta diretamente a sua compreensão e, assim, a extensão de seu uso.

De acordo com Leviton e Hughes (1981), a utilização pode ser prejudicada se os incentivos para ler os relatórios de avaliação são baixos. Como exemplo, os autores citam os jargões e vocabulários específicos utilizados pelos avaliadores como um dificultador do entendimento dos resultados das avaliações pelos usuários, uma vez que, para a maioria dos gestores e políticos, a avaliação ainda constitui uma ferramenta incomum.

O estilo dos administradores dos programas sociais também afeta diretamente a utilização das avaliações, como afirmado por Cox *apud* Leviton e Hughes (1981) a partir da pesquisa de Mintzberg (1973) sobre as atividades gerenciais. As implicações do estudo alertaram que: 1) somente os resultados relevantes às necessidades dos gerentes receberão atenção; 2) é improvável que seja realizado um exame crítico da qualidade dos resultados; 3) a comunicação verbal facilitará a utilização, uma vez que os gerentes preferem ser continuamente informados. Além disso, como afirmado por Leviton e Hughes (1981), os administradores preferem informações qualitativas a informações quantitativas, o que requer a combinação de métodos mistos.

A diferença de estilos entre os administradores e os pesquisadores também poderão afetar a utilização (LEVITON; HUGHES, 1981). Como dito pelos autores, existe pouco contato entre os administradores e a comunidade acadêmica, dado que seus valores são distintos.

Quanto à variável credibilidade, Leviton e Hughes (1981) afirmam que os administradores utilizam as avaliações também para aumentar a certeza sobre conhecimentos e decisões já existentes. Nesse caso, como dito por Patton *et. al.* (1977), avaliações são realizadas para corroborar informações sobre as quais não se tem certeza ou, até mesmo, justificar ações. Assim, a percepção dos administradores sobre a validade de uma pesquisa interfere

fortemente em sua avaliação das mesmas – os administradores que preferem pesquisa ao contrário da intuição são potenciais usuários das avaliações.

A credibilidade do avaliador (produtor de uma avaliação) é também um fator de extrema importância (LEVITON; HUGHES, 1981). Guba (1975) atenta para o fato de que muitos avaliadores são cooptados por interesses políticos e induzidos a escrever o que o cliente quer escutar. Nesse sentido, o risco de cooptação de avaliadores e da supressão de informações aparece como um fator que prejudica fortemente a utilização.

Embora alguns autores tenham apontado que a qualidade metodológica de uma avaliação favoreça fortemente a sua utilização, a qualidade não afeta diretamente a utilização, como afirmado por Patton et al. (1977). Uma metodologia ausente de questionamentos, contudo, e a confiança nos dados apresentados, por sua vez, tornam-se essenciais para o convencimento e influenciam diretamente a utilização (LEVITON; HUGHES, 1981). No estudo de Weiss e Bucuvalas (1977), a qualidade da pesquisa estava relacionada a dois tipos diferentes de utilização: mobilizar apoio para um posicionamento e modificar maneiras de pensar sobre uma questão.

A última variável apresentada por Leviton e Hughes (1981) considera a participação dos usuários e *advocacy* (aqui entendida como apoio). De acordo com os autores, a relação entre avaliações e a *advocacy* está inserida no contexto político em que as avaliações são conduzidas: o apoio político e compromisso de um indivíduo-chave com a avaliação é fundamental para a utilização dos seus resultados. Essa variável, no entanto, é limitada pelas mudanças súbitas na política e pela rotatividade presente nos cargos do alto escalão governamental.

O quadro 5 resume as variáveis descritas por Leviton e Hughes (1981). O sinal positivo (+) posicionado à esquerda de uma variável aumenta a utilização e um sinal negativo (-) diminui a utilização. Um sinal de mais ou menos (+/-) indica que a variável interage com outras variáveis, aumentando a utilização em alguns casos, diminuindo em outros e ainda não tendo efeito em outros.

Quadro 7 – Fatores relacionados à utilização das avaliações para Leviton e Hughes (1981)

(continua)

Variáveis		
Relevância	Avaliações direcionadas à necessidade dos clientes	+ Avaliações direcionadas à necessidade dos clientes
	Necessidades dos tomadores de decisão	+ Recursos na alocação da informação
		+ Informação implementável
		+ Informação sobre a efetividade geral do programa
	Necessidade dos gestores dos programas	+ Informação implementável
		+ Informações sobre a efetividade de elementos do programa
		- Informação sobre a efetividade geral do programa tem menor prioridade
	Tempestividade	+/- Provável interação com o tipo de uso
		+/- Interação com a fase do ciclo político
		+ Tempo de espera para apreciação de evidências
Comunicação	Comunicação direta dos usuários sobre as necessidades para a avaliação	+ Comunicação direta dos usuários sobre as necessidades para a avaliação
	Comunicação por intermédio da hierarquia da burocracia	- Comunicação por intermédio da hierarquia da burocracia
	Disseminação	+ Disseminação direta
		- Informação esquecida nos resumos executivos das avaliações
		- Distorções ou declarações incorretas nos resumos
		- Informações negativas suprimidas
		+ Rede de disseminação
		- Disseminação entre agências
		+ *colégios invisíveis* de especialistas
Processamento de informações	Reconhecimento da relevância	+ Avaliações planejadas para responder a questões específicas
		- Objetivos mal delineados
		- Relevância das avaliações locais para o governo Federal
		+/- Informações imprevistas - provável interação com habilidades do usuário
	Apresentação das informações	+ Apresentação clara
		- Jargões que aumentam as dificuldades
	Processamento de informações de acordo com o estilo dos administradores	- Atenção a informação relevante apenas
		- Pouco controle da qualidade das informações
		+ Comunicação verbal
		- Falta de familiaridade com os métodos utilizados
		+/- Preferência por informações qualitativas
		- Diferença entre administradores e estilos acadêmicos
Credibilidade	Comparação com outras informações	- Contradição entre pesquisa e outras informações
		- Contradição entre com intuição e expectativas dos usuários
		+ Informações colaborativa
		+ Uso da pesquisa em agregados
	Preconceitos dos usuários	+ Opinião elevada sobre pesquisa
		+ Confiança na pesquisa em detrimento à intuição
		+ Usuário com formação médica
		- Usuário fom formação em direito – uso persuasivo
		+ Usuário está temporariamente no governo

Quadro 8 – Fatores relacionados à utilização das avaliações para Leviton e Hughes (1981)

(conclusão)

	Credibilidade do avaliador	- Provável efeito de cooptação suspeita
		- provável efeito sobre o uso de informações exterior à organização
	Qualidade	- Qualidade baixa detectada
		- Ausência de informação necessária
		+/- Interação com diferentes tipos de uso
		+ Metodologia não é facilmente contra-argumentada
Participação e *advocacy* dos usuários	Comprometimento com a avaliação	+ Elevado interesse e comprometimento do usuário
		+ Defesa da informação da avaliação
		+/- Provável interação do comprometimento do usuário com a mudança na política
		+/- Provável interação do comprometimento do usuário e a rotatividade dos mesmos
	Defesa ao programa e políticas	+/- Interação com os resultados diretos
		+ Avaliações menos visíveis
		+ menor quantidade de deliberadores
		+/- Provável interação entre os resultados da avaliação e sua tempestividade

Fonte – Leviton e Hughes, 1981, adaptado pela autora da dissertação.

Cousins e Leithwood (1986) revisaram as pesquisas sobre utilidade das avaliações no período de 1971 a 1986. Posteriormente, outros autores dedicaram-se a trabalho semelhante, porém em períodos diferenciados e com outro conjunto de teóricos, o que culminou na discussão e surgimento de informações sobre outras variáveis que também poderiam influenciar a utilidade. O trabalho de Shulha e Cousins (1997) é um deles.

3.3.2 Shulha e Cousins (1997): o papel da política e das estruturas e processos organizacionais

De acordo com Shulha e Cousins (1997), em 1996, a pesquisa de ambos os autores foi responsável não somente por marcar o 10º aniversário da Associação Americana de Avaliação (American Evaluation Association – AEA), como também prestigiar uma década de atividades de estudos e trabalhos interessados na natureza, causas e consequências da utilização das avaliações. Nesse momento, na seção nomeada *The past, present and future of evaluation use: the continuing evolution of evaluation practice and theory*, apresentou-se a revisão dos trabalhos de utilização, no período de 1986 e 1996, feito pelos autores.

A pesquisa de utilização após 1986, para Shulha e Cousins (1997), já havia conseguido provar a complexidade do tema do uso das avaliações, dos programas e dos contextos políticos em que a avaliação ocorreu. Nesse ponto de desenvolvimento da pesquisa, os teóricos do campo

já haviam reconhecido que o *projeto de avaliações úteis* exigia muito mais do que simplesmente a combinação de perguntas dos usuários das avaliações sobre os métodos adequados, abordagens ou modelos de avaliação. Cada vez mais os avaliadores estavam falando sobre a necessidade de identificar o tipo de uso apropriado aos diferentes *stakeholders* envolvidos nas avaliações e a confecção e desenho de avaliações cada vez mais customizadas e planejadas para a potencialização do uso.

Shulha e Cousins (1997) argumentam que os estudos que tentaram catalogar as variáveis que influenciariam a utilização, no período de 1986 e 1996, podem ser divididos em três fases principais. A primeira delas leva em conta o papel do contexto em que as avaliações estão inseridas. Alguns teóricos de destaque, como Guba e Lincoln (1989), declaram que problemas comuns podem comprometer a utilização das avaliações. Dentre esses problemas, os autores citam a hegemonia do paradigma positivista, a incapacidade em acomodar o pluralismo de atores envolvidos nos programas públicos e a desconsideração do contexto, das circunstâncias e propósito que envolve a implantação de programas. Os autores desafiaram a confecção de avaliações tradicionais e propuseram um método construtivista para a sua realização, uma vez que, na percepção dos autores, a construção de conhecimento é parte integrante contexto em que ele é produzido.

Uma segunda perspectiva trazida por Shulha e Cousins (1997) que marcou a pesquisa de utilização de avaliações enfatiza que a atividade política também estaria relacionada ao contexto. Nesse veio, a tomada de decisões alinhada às recomendações de uma avaliação não dependeria somente de um único ator, mas de outros grupos de interesse e circunstâncias do próprio setor público.

Uma terceira abordagem dos estudos do uso de avaliações de políticas relaciona-se à estrutura e aos processos organizacionais em que as avaliações estão inseridas (SHULHA; COUSINS, 1997). Por meio dessa perspectiva, consideram-se os canais de comunicação da organização e a burocracia de processos que, em menor ou maior grau, influenciam, diretamente, a transmissão das informações geradas pelas avaliações e a inserção de mudanças no nível organizacional.

Outros pesquisadores, igualmente preocupados com fatores contextuais, focaram o papel da interação entre avaliadores e o contexto do programa em questão. Um dos marcos mais

importantes para o campo, de acordo com Shulha e Cousins (1997), está inserido no debate sobre o papel dos avaliadores em assegurar a utilização das avaliações entre dois autores centrais: Carol Weiss e Michael Patton.

O debate entre Carol Weiss e Michael Patton (*Weiss-Patton debate*) contribuiu de forma significativa para o desenvolvimento da teoria de utilização, como afirmado por Shulha e Cousins (1997). As discussões entre os autores, iniciadadas em 1987 na AEA, posteriormente se traduziram em uma série de artigos do *American Journal of Evaluation* e se tornou um marco no campo. O ponto de partida do debate foi o clássico trabalho de Weiss (1988) *Is anybody there? Does anybody care?*, cujo foco central aborda a fraca utilização dos resultados das avaliações em mudanças concretas nos programas e ações governamentais, ou seja, a ausência de tomada de decisões advinda de estudos avaliativos contratados para esse fim. A constatação da autora baseava-se em cinco argumentos centrais: 1) natureza política dos processos de decisão; 2) as avaliações não são capazes de dar respostas pragmáticas alinhadas aos anseios dos tomadores de decisão; 3) as decisões não são tomadas com base em um processo racional; 4) muitas vezes não se conhece o tipo de informação necessária para se tomar decisões; 5) interesses diversos protegem a continuidade do *status quo* do programa avaliado.

De acordo com Shulha e Cousins (1998), Michael Patton que, no mesmo momento, discordou dos argumentos da autora, ressaltou que, embora as conclusões de Weiss (1988) fossem válidas para análise do processo decisório, em se tratando de avaliações, outros argumentos deveriam ser levados em conta. Na visão do autor, a perspectiva de Weiss (1988) seria limitada, vez que a autora se restringia à utilização das avaliações unicamente direcionadas à tomada de decisões.

O debate entre os autores Carol Weiss e Michael Patton continuou de forma mais aprofundada, centrado na questão do papel dos avaliadores (SHULHA; COUSINS, 1997). Para Carol Weiss, como os processos decisórios no setor público são caracterizados pela irracionalidade e pluralismo, seria papel dos avaliadores encontrar a melhor maneira – método e abordagem avaliativa – de auxiliar os usuários das avaliações. Para Michael Patton, os avaliadores teriam um papel mais ativo na promoção e no cultivo do uso, além de auxiliar os decisores quanto à abordagem da avaliação. Enquanto o impasse não era resolvido, a

importância do contexto em que os avaliadores estavam inseridos era cada vez mais valorizado pelos teóricos do campo, visto por Shulha e Cousins (1997).

Expandindo a visão do contexto para os demais fatores que impactam a utilização das avaliações, outros autores observaram como a participação de *stakeholders* no processo de avaliação é essencial para alcançar confiança na qualidade da informação gerada. Conforme afirmado por Patton (2008), como a utilização das avaliações é inerente aos valores e objetivos daqueles que farão uso das informações por elas produzidas, não existe uma definição absoluta e ideal para *utilização*.

Em decorrência disso, Patton (2008) enfatiza a importância da participação dos usuários interessados nos resultados da avaliação durante todo o seu processo: *utilization-focused evaluation*. As avaliações com foco no uso são avaliações feitas para e com usuários específicos almejados. Para Patton (2008), as avaliações com foco no uso devem ser julgadas pela sua utilidade real; assim, ao desenvolver uma avaliação, os avaliadores devem planejar, do desenho à conclusão, do início ao fim, todos os processos e fatores que irão afetar o uso. Nesse sentido, de acordo com Patton (2008), a ideia central das *utilization-focused evaluation* é que os avaliadores *cerquem* todas as possibilidades que possam restringir o uso de uma avaliação. Tudo deve ser pensado e planejado tendo-se em vista a utilização da mesma por clientes específicos, ou seja, "[...] uso pretendido por clientes pretendidos"(PATTON, 2008, p. 37).

Após Cousins e Leithwood (1986) e Shulha e Cousins (1997), Johnson *et al.* (2009) revisam as pesquisas sobre utilidade das avaliações no período de 1986 a 2005, como será apresentado na seção seguinte.

3.3.3 Johnson *et al.* (2009): a importância do envolvimento dos *stakeholders*

Johnson et al. (2009) conduziram uma pesquisa sobre a utilização das avaliações em um período de 25 anos, compreendido entre 1986 a 2005. A pesquisa consistiu na revisão de periódicos de avaliação – *American Journal of Evaluation, Canadian Journal of Program Evaluation, Evaluation, Evaluation Practice, Evaluation and Program Planning, Evaluation Review, New Directions for Evaluation, and Studies in Educational Evaluation* – dissertações e capítulos de livros, a partir de termos como *evaluation use*, *evaluation utilization* e

evaluation influence. Após a revisão da pesquisa, a equipe de pesquisadores identificou 98 publicações que obedeciam aos critérios de análise e, destes, apenas 41 estudos foram considerados pelos autores como adequados.

Como sugerido por Shulha e Cousins (1997), mudanças nas conceptualizações sobre o uso das avaliações ocorreram e, assim, novas categorias apareceram. No estudo de Johnson et al. (2009), uma nova característica – competência do avaliador – foi adicionada à categoria de implementação das avaliações. O estudo também foi responsável por criar uma nova categoria completa: envolvimento dos *stakeholders* para acomodar 25 estudos examinados que não correspondiam aos critérios observados por Cousins e Leithwood (1986).

Esse arcabouço conceitual, longe de ser definitivo, ajudará a analisar como as avaliações tendem a modificar as instituições públicas e a dificuldade existente em utilizar, de forma apropriada, seus resultados. É possível perceber, à primeira vista, uma necessidade de integração dessas abordagens para analisar a utilização das avaliações face ao ambiente e capacidade institucional de organizações públicas.

A característica de competência do avaliador foi adicionada, de acordo com os autores, pois seis estudos apontaram que os profissionais da avaliação desempenham um importante papel na condução da sua utilização, por diferentes razões. Embora o estudo de Cousins e Leithwood (1986) apontasse a reputação dos avaliadores na característica de credibilidade, a definição não compreendia a natureza influenciadora do avaliador, sua competência ou liderança no momento de utilizar os resultados de uma avaliação.

A nova categoria envolvimento dos *stakeholders* foi adicionada devido a estudos mais recentes como apontado por Johnson *et al.* (2009). Dos estudos analisados, mais da metade (23 de 41) apresentaram alguma preocupação com o envolvimento de atores. A nova categoria tem o objetivo de refletir os estudos avaliatórios mais participativos.

Quadro 9 – Fatores relacionados à utilização das avaliações para Johnson et al (2009)

(continua)

Categoria	Características	Descrição
Implementação da avaliação	Qualidade da avaliação	Características do processo de avaliação, sofisticação de métodos, rigor, tipo da avaliação
	Credibilidade	Objetividade, credibilidade e adequação do processo de avaliação e / ou das atividades do avaliador
	Relevância	Extensão da informação fornecida da avaliação que é relevante para o tomador de decisão e a localização organizacional do avaliador
	Qualidade da comunicação	Clareza e frequência de comunicação dos resultados, amplitude da divulgação. Inclui também o tipo de recomendações do relatório e do processo de comunicação entre avaliadores e clientes
	Resultados	Natureza dos resultados das avaliações (positivos ou negativos) extensão de coerência com as expectativas do público, o valor dos resultados para tomada de decisão
	Tempestividade	Tempestividade da avaliação em um contexto ampliado; tempestividade de comunicação quando a avaliação é concluída; tempestividade de disseminação aos tomadores de decisão
	Competência do avaliador	Características pessoais do avaliador, competências culturais, estilo de liderança do avaliador
Características da decisão ou do *policy setting*	Necessidade da informação	Necessidade de informação dos públicos das avaliações ; tipos de informações; características do público, com diferentes necessidades de informação
	Características da decisão	A singularidade da decisão ou do problema que a avaliação tenta resolver; tipo de decisão a ser tomada e a novidade da área do programa
	Clima político	Orientação política da(s) pessoa(s) que encomendou as avaliações; grau em que a decisão é dependente de patrocinadores externos, rivalidades internas, constrangimentos de orçamento e lutas de poder
	Informação competitiva	Informações relacionadas ao tema da avaliação e à disposição dos interessados externos ao processo de avaliação
	Características pessoais do usuário	Características do usuário de avaliação, por exemplo, papel organizacional do tomador de decisão, estilo de processamento de informação, características sociais, e assim por diante
	Comprometimento e receptividade à avaliação	Atitudes e comprometimento dos usuários das avaliações com a condução das mesmas; grau de resistência organizacional quanto à avaliação.
Envolvimento dos *stakeholders*	Envolvimento com comprometimento ou receptividade à avaliação	O envolvimento de atores cria comprometimento e receptividade com a avaliação
	Envolvimento com comunicação de qualidade	O envolvimento de atores promove o aprimoramento da comunicação
	Envolvimento direto do *stakeholder*	A relação direta entre envolvimento de atores e o uso da avaliação
	Envolvimento com credibilidade	Envolvimento dos atores levam ao aumento da credibilidade do processo de avaliação
	Envolvimento com resultados	Envolvimento dos atores no entendimento dos resultados das avaliações.
	Envolvimento com relevância	Participação de atores integrados com as preocupações organizacionais do desenho da avaliação
	Envolvimento com características pessoais	Envolvimento de atores de diferentes níveis organizacionais

Quadro 10 – Fatores relacionados à utilização das avaliações para Johnson et al (2009)

(conclusão)

	Envolvimento com características de decisão	O envolvimento de uma série de atores inseridos em configurações diferentes, dependendo das características das decisões que precisam ser realizadas
	Envolvimento com necessidade de informações	O envolvimento dos atores facilita a introdução de suas necessidades por informação

Fonte – Johnson *et al.*, 2009, adaptado pela autora da dissertação.

Após a discussão do tema da utilização das avaliações, torna-se possível aproximar o tema das avaliações da experiência nacional e estadual. O capítulo que se segue, contextualiza o tema das avaliações no setor público brasileiro e, posteriormente, no governo do Estado de Minas Gerais, foco do estudo exploratório deste trabalho.

4 AS AVALIAÇÕES DE PROGRAMAS NO SETOR PÚBLICO

A avaliação de políticas sociais constitui um campo articular da pesquisa social que se expande a partir da década de 1960 nos Estado Unidos, no bojo dos programas do *welfare state* desenvolvidos naquele país. Buscava-se, com isso, encontrar metodologias que permitissem avaliar a efetividade das intervenções estatais na área social, por meio da construção de métodos pautados em critérios técnicos que visavam respaldar as ações de governo e indicar alternativas com maior benefício social, corrigindo os rumos dos programas.

No Brasil, ao contrário, não é possível encontrar o mesmo movimento. A pesquisa de avaliação de políticas sociais, no contexto nacional, expande-se apenas nos anos 1990 com o processo de reforma da administração pública gerencial (CALMON; GUSSO, 2002), seguindo mais as pressões de organismos internacionais (RIBEIRO, 2009) do que, diretamente, a melhoria da qualidade da provisão de serviços públicos e políticas mais efetivas. Dessa forma, a aplicação de estudos avaliatórios sistemáticos nos programas sociais brasileiros constitui uma prática ainda recente no contexto nacional, alinhada mais à gestão do que ao seu impacto sobre a necessidade de formulação de políticas sociais mais equânimes, como direito inalienável do cidadão na construção da democracia, como visto na primeira seção deste capítulo.

O protagonismo estadual quanto às práticas de avaliação, seguindo o caminho brasileiro, também se iniciou de forma tardia de acordo com as reformas gerenciais que se instalaram nos diferentes entes da federação. O estado de Minas Gerais, destaque, a partir de 2003, em práticas de modernização da administração pública, remonta uma experiência em monitoramento e avaliação. Na segunda seção deste capítulo apresentam-se a trajetória de Minas Gerais e, em um segundo momento, as avaliações realizadas no período de 2007 a 2010.

4.1 Antecedentes: reformas da administração pública e o contexto brasileiro

O processo de institucionalização da avaliação na América Latina ocorreu tardiamente. A perspectiva prevalecente foi a de instrumentalização da pesquisa avaliativa para o sucesso da

reforma do Estado. E, mesmo acompanhando as atividades referentes à modernização do governo, a função avaliação no Brasil é ainda bastante incipiente, com institucionalização precária, como comenta Faria (2005). De acordo com Lobo (2001), a administração pública brasileira, à semelhança de outros países latino-americanos, não desenvolveu ao longo do tempo uma cultura de efetividade em ralação ao gasto público com os programas sociais.

No Brasil, a adoção de práticas de gestão pública por resultados foi prejudicada nas últimas décadas por duas razões (CALMON; GUSSO, 2002): a primeira reside no fato de que o combate à inflação acelerada, a negociação da dívida externa e sucessivos intentos de retomar o crescimento tornaram-se temas quase exclusivos da agenda política e econômica do País; a segunda razão consiste que, no ambiente político-econômico reinante antes da década de 90, planejar os médio e longo prazos parecia não apenas pouco viável, como, impertinente politicamente.

Lobo (2001) ressalta que a predominância de aspectos autoritários na administração pública brasileira também se coloca como um dos obstáculos para se implantar uma cultura de avaliação de programas governamentais. Para a autora, há de convir que o ambiente político nos quais os programas são formulados, associado ao constrangimento das autoridades e gestores públicos frente às perspectiva de avaliação, inibe qualquer iniciativa que tenha como objetivo a verificação técnica de resultados reais das ações do governo. Além disso, o temor quanto a resultados inexpressivos ou inexistentes também dificulta a penetração dos estudos avaliatorios. De acordo com Arretche (1998), em sociedades em que o mercado eleitoral tem peso, o impacto social de uma política tende a ter menor importância no processo decisório que seu impacto sobre a opinião pública ou seu impacto eleitoral, como a autora aborda nesse trecho destacado abaixo.

> A avaliação de eficiência, eficácia e efetividade das políticas tende a ser apenas um dos elementos – e possivelmente de muito menor importância – na decisão pela adoção, reformulação ou supressão de um programa público, dado que as razoes do mercado eleitoral tem forte influência no mercado decisório (ARRETCHE, 1998, p. 36).

De acordo com Garcia (2001), quatro razões podem ser citadas para justificar a demanda por avaliações das políticas sociais. Em primeiro lugar, a crise fiscal brasileira diminuiu a capacidade de gastos dos governos e expôs os problemas das finanças públicas. Em segundo, o aprofundamento da democracia, responsável por trazer novos atores sociais e reivindicações

aos governos. Um terceiro ponto estaria relacionado à longa crise econômica responsável por aumentar a desigualdade social e a busca pelos programas sociais do governo. Por último, pode ser citada a desestruturação da função de planejamento, que deixou os governantes sem elementos de informação.

Costa e Castanhar (2003) afirmam que, historicamente, na administração pública brasileira não há a preocupação de avaliar programas públicos, sejam esses gerais, sejam projetos sociais. A pesquisa de avaliações começa a entrar em pauta no setor público brasileiro somente na década de 1980, no contexto de transição política e da necessidade de instituir programas sociais mais equânimes, segundo Castro (1989). Prova disso são os estudos diagnósticos que analisaram a perversidade do sistema social brasileiro consolidado pós-1964, indicando a precisão por programas que combatessem o sistema de desigualdades presente na estrutura social brasileira (CASTRO, 1989).

A Constituição Federal de 1988, em seu artigo 165, inseriu o Plano Plurianual (PPA) como forma de balizar os programas e ações governamentais com as previsões financeiras em quatro anos. No mesmo artigo, criou-se também a Lei de Diretrizes Orçamentárias (LDO), que estabelece orientações para a elaboração do orçamento anual. E, por fim, estabeleceu-se a unificação dos orçamentos anuais (fiscal, seguridade, investimento das estatais) que, em 1998, a ser chamada de Lei do Orçamento Anual (LOA) – responsável por estimar as receitas que o governo arrecada durante o ano e fixar os gastos realizados com tais recursos. A Lei de Diretrizes Orçamentárias e a Lei do Orçamento Anual têm de ser compatíveis com o PPA, bem como todos os planos e programas nacionais, regionais e setoriais previstos na Constituição, significando um importante processo de planejamento e gestão pública.

Entretanto, persistiam algumas dificuldades para que tais instrumentos pudessem prosperar como parte integrante dos processos de gestão, já que pouco se sabia da origem e rotina de implementação dos programas, o que tornava penoso identificar os objetivos que seguiam e, assim, estabelecer critérios consistentes para avaliá-los. O primeiro PPA, elaborado para o período 1991-1995, teve como objetivo somente cumprir a determinação constitucional, sem qualquer efeito no estabelecimento de metas econômicas de governo (CALMON; GUSSO, 2002).

O Plano Real, implantado em 1994, trouxe a estabilidade da moeda e as condições para a retomada do planejamento. Após a estabilização econômica, foi lançado o Plano Diretor da Reforma do Estado, baseado nos princípios da nova gestão pública e da reforma administrativa de 1995, com a finalidade de reestruturar a administração pública conferindo-lhe flexibilidade e responsabilização. Segundo Nader (2005), nesse período, intensificou-se o movimento na direção dos princípios e valores da gestão empreendedora no setor público.

> O Plano Diretor focava a flexibilização da gestão pública por meio da transformação das estruturas jurídicas das organizações, a descentralização das ações governamentais e a responsabilização dos gestores públicos pelos resultados das ações institucionais, e disseminava os princípios da administração pública gerencial. O Plano Diretor não foi implementado na íntegra, mas deixou raízes no processo de incorporação de novos conceitos e princípios na administração pública [...]. (NADER, 2005, p.1)

O PPA 1996-1999 trouxe novos conceitos como os eixos estratégicos de integração do *Programa Brasil em Ação*, entre eles, transportes, energia, telecomunicações e novas tecnologias. Contudo, a decisão de transformar o plano em instrumento de gestão, inaugurando a administração pública para resultados, surgiu com o Plano Plurianual 2000-2003, denominado avança Brasil (GARCES; SILVEIRA, 2002). Autores como Almeida (2006) compartilham do mesmo pensamento.

> O PPA seguinte, válido para o período 2000-2003, inovou ao introduzir na Administração Pública Federal a gestão por resultados, com a adoção de programas como unidade de gestão, a integração entre o plano, o orçamento e a gestão e o fortalecimento do conteúdo estratégico por meio do estudo dos eixos (ALMEIDA, 2006, p. 222).

Outro instrumento de gestão, concomitante e complementar ao PPA, foi criado com a Lei de Responsabilidade Fiscal (LRF), um novo sistema de controle das despesas públicas, estabelecendo limites para os pagamentos de pessoal, para a dívida pública, para os gastos correntes e despesas extraordinárias. De acordo com Garces e Silveira (2002), a lei permitiu um planejamento fiscal de médio prazo importante para a orientação do gasto futuro, elemento indispensável para a credibilidade do planejamento.

A avaliação dos programas sociais no Brasil ganha força somente nos anos 1990, motivada por muitos fatores, dentre os quais, a crise de financiamento do setor público e as iniciativas de organismos internacionais de financiamento, tais como o Banco Mundial (BIRD) e o Banco Interamericano de Desenvolvimento (BID), no sentido de instrumentalizar países em

desenvolvimento para gestão pública mais eficiente e efetiva (RIBEIRO, 2009). Calmon e Gusso (2002, p.7), ao apresentarem a experiência recente do governo federal em avaliação, também compartilham do argumento, afirmando que,

> [...] em meados da década de 1990, ganhou destaque o reconhecimento da necessidade de o governo [brasileiro] dispor de processos de monitoramento e avaliação que lhe permitissem controlar e articular suas ações e ganhar legitimidade por meio de eficaz comunicação de seus êxitos à sociedade.

Nesse sentido, a preocupação com a efetividade das políticas sociais estaria mais relacionada à reforma do setor público e aos pressupostos da gestão eficiente do que, de fato, um sistema consolidado de proteção social e a necessidade de desenvolver políticas que atendam às necessidades da população. A reforma administrativa institucionalizada a partir de 1995 no Brasil havia, assim, estabelecido o marco institucional para a promoção de reformas administrativas nos demais entes da federaçação, um protagonismo estadual em torno de princípios da administração gerencial e da cultura de resultados começa a surgir. Em Minas Gerais o planejamento estratégico e a gestão pública por resultados tornaram-se aspectos marcantes do processo de implementação da reforma administrativa (CORRÊA, 2007).

No entanto, embora muitas questões tenham avançado no campo das políticas sociais, ainda impera um espaço significativo de indeterminação e irresponsabilidade. Ainda prevalecem as decisões individualistas, incertas e ignorantes na implementação dos programas sociais brasileiras, apesar da carência por programas sérios que tragam algum retorno à população. Na seção seguinte, apresenta-se o choque de gestão do governo de Minas Gerais de forma a ilustrar a abordagem teórica até aqui elaborada, particularmente referente ao modelo de gestão por resultados.

4.2 O Choque de Gestão em Minas Gerais

Como indicado no estudo *Minas Gerais do século XXI*, em 2003, o estado de Minas Gerais atravessava um contexto de agravamento do quadro fiscal e graves problemas de desempenho da administração pública (VILHENA, 2006). Em decorrência desse quadro, buscou-se, a partir do programa Choque de Gestão, uma resposta à crise fiscal e administrativa num esforço de alcançar o equilíbrio das contas públicas e modernizar o aparato institucional do Estado.

De acordo com Anastasia (2006), o Choque de Gestão, linha de conduta principal do programa governamental do Estado, foi um conjunto de medidas de impacto, implantadas entre 2004 e 2007, para modificar o padrão de comportamento da administração estadual. Nesse âmbito, como uma de suas medidas, foi implantada uma abordagem dual para o planejamento. O duplo planejamento consiste, em síntese, em dois planejamentos, coerentes entre si, sendo um de curto prazo, correspondente às medidas imediatas do Choque de Gestão, e outro de médio e longo prazos. A formulação da estratégia dual do Estado sustentou-se, portanto, em dois pilares: o equilíbrio fiscal, estratégia do presente, e uma agenda de desenvolvimento futura de médio e longo prazos (GUIMARÃES; ALMEIDA, 2006).

Um ponto de partida inequívoco para a implantação do sistema de monitoramento e avaliação mineiro foi o fortalecimento das ferramentas de planejamento já existentes[1], de modo que pudessem constituir uma plataforma para integrar a orientação por resultados no governo (BUSJEET, 2010). Como instrumento de planejamento de longo prazo, foi aprovado pela Assembleia Legislativa de Minas Gerais (ALMG), em 2004, o Plano Mineiro de Desenvolvimento Integrado (PMDI), apresentando-se como o plano estratégico norteador da ação governamental do estado de Minas Gerais (GUIMARÃES; CAMPOS, 2008). O PMDI instituiu a visão de Minas como o melhor estado do País para se viver e investir em 2020, mediante melhora contínua do Índice de Desenvolvimento Humano (IDH).

Após a consecução da estratégia de longo prazo – PMDI – foi determinado um portifólio de 30 projetos estruturadores, projetos que continham o detalhamento das ações gerenciais prioritárias para o atingimento dos resultados previstos, são gerenciados com intensidade e têm fluxo garantido de recursos orçamentários. Tais projetos estão contidos no Plano Purianual de Ação Governamental (PPAG), um desdobramento do PMDI para um período de quatro anos, que detalha os programas e ações governamentais, com as previsões financeiras e físicas (GUIMARÃES; ALMEIDA, 2006). A carteira de projetos estruturadores foi submetida à Superintendência Central de Gestão Estratégica de Recursos e Ações do Estado (SCGERAES), ainda em 2004. Adquirindo a função de escritório de projetos, o SCGERAES apresentou, como principal atribuição, o monitoramento intensivo da carteira recém-criada, ou

[1] Apesar de os planos de longo prazo constituírem uma imposição da Constituição Mineira desde 1989, de acordo com Busjeet (2010), mostrava que o exercício não fora usado de forma plena e significativa como forma de planejamento estratégico para influenciar a formulação de políticas.

seja, seria responsável por gerenciar os projetos de forma diferenciada, estabelecendo os planos, as metas e acompanhando o desempenho em relação ao planejado (BECHELAINE, 2009).

A adoção da contratualização de resultados, instituída pelo Acordo de Resultados, e do monitoramento da agenda setorial, também fruto do período, foram promovidos pela Superintendência Central de Modernização Institucional (SUMIN). De acordo com Carneiro e Barbosa (2007, p. 2), "[...] o Acordo é um instrumento de pactuação de resultados, que se dá mediante a negociação entre dirigentes dos órgãos e entidades do Poder Executivo, bem como dos gerentes das unidades administrativas e autoridades que sobre eles tenham poder hierárquico ou de supervisão". Essa ferramenta gerencial celebra um *contrato* que representa um compromisso (negociado entre as partes) que regula as relações (via de regra) entre o núcleo estratégico (formulador de políticas) e as entidades descentralizadas (executoras) (DUARTE, 2006), ou seja, que coordena as metas e resultados, visando a melhoria do desempenho operacional das organizações partícipes.

Enquanto o desenvolvimento do escritório de projetos iniciou o elo entre programas governamentais e produtos desses programas, a contratualização de resultados principiou o elo entre programas e organizações. Destaca-se, assim, o papel do GERAES como *a base* da gestão pública por resultados no Estado, e a relevância dos acordos de resultados como instrumentos que iniciaram a gestão por resultados mineira. Como completa Corrêa (2007, p. 498),

> [...] a gestão pública por resultados foi estruturada por meio de procedimentos de contratualização em um instrumento de avaliação de desempenho chamado de acordo de resultados. Por meio do acordo de resultados o governo de Minas Gerais, mais especificamente centralizado na Secretaria de Estado de Planejamento e Gestão, passou a garantir que as demais secretarias de estado estariam buscando os resultados para o desenvolvimento do Estado.

Em 2007, o PMDI foi reestruturado para o horizonte 2007-2023, apresentando-se na Lei 17.007, de 28 de setembro de 2007. O novo plano buscou, por meio da mesma visão de futuro *Tornar Minas o Melhor Estado para se Viver*, focar os resultados finalísticos das áreas de resultado. De acordo com o próprio documento, as áreas de resultado são *áreas focais* de concentração dos melhores esforços e recursos visando transformações e melhorias desejadas. Cada área de resultados seria alvo da intervenção de um grupo de projetos estruturadores,

reunidos em função de sua sinergia, de forma que seus resultados seriam monitorados por um grupo de indicadores finalísticos previstos no PMDI (MINAS GERAIS, PMDI, 2007). Já os resultados finalísticos estabelecem os indicadores e metas, sejam elas para 2007, 2011, seja para 2023, para cada uma das áreas de resultados. Acima das áreas de resultados, foi acrescida a estratégia denominada Estado para Resultados, cujo foco é assim descrito "A estratégia do Estado para Resultados pretende integrar, em Minas, a análise dos resultados para a sociedade *vis-à-vis* o ônus determinado pela arrecadação de receitas públicas, com o intuito de se obter a máxima eficiência alocativa do orçamento do Estado" (MINAS GERAIS, PMDI, 2007, p. 19).

Comparando-se o PMDI 2003-2020 e a versão de 2007, é importante notar que o primeiro não estabelecia metas a serem atingidas e se aproximava mais da carteira de projetos a ser executada, enquanto o segundo, além de incorporar metas, deixa a especificação dos projetos estruturadores para o PPAG. Dessa maneira, o PMDI 2007-2023, ao discutir os objetivos e resultados que o Estado almeja alcançar, passou a focar o resultado, ao invés da estrutura organizacional (GUIMARÃES; CAMPOS, 2008). Além disso, tratando-se de um modelo de gestão por resultados, uma questão fundamental é a tradução de objetivos em metas que, por sua vez, são mensuradas por meio de medidas dos resultados que deverão ser alcançados, os indicadores. O PMDI 2007-2023 operacionaliza esses objetivos por meio de indicadores finalísticos que traduzem as condições econômicas, sociais, ambientais, etc. Por exemplo, o objetivo de *ampliar o acesso ao saneamento básico* pode ser operacionalizado por meio da apuração do indicador finalístico *percentual de domicílios com acesso à rede coletora de esgoto*. Assim, para tanto, além da implantação do acordo de resultados, considerado como o princípio da gestão por resultados no governo mineiro, as mudanças instituídas pelo PMDI 2007-2023, na segunda geração de políticas, também contribuíram para reforçar o modelo de resultados, na medida em que operacionalizaram metas e indicadores de desempenho para o Estado.

Para tanto, além da implantação do acordo de resultados – considerado como o princípio da gestão por resultados no governo mineiro, e das mudanças instituídas pelo PMDI 2007-2023, a operacionalização de metas e de indicadores de desempenho para o estado, bem como implementação de avaliações dos projetos estruturadores na segunda geração de políticas, foco da próxima seção, também contribuíram para reforçar o modelo de resultados.

4.3 A gestão por resultados em Minas Gerais

Outra mudança instituída no quadriênio 2007-2011 foi a criação do Programa Estado para Resultados (EpR), uma das unidades protagonistas da agenda de avaliações de políticas públicas do governo. O EpR, caracterizado como *guardião do PMDI* (BUSJEET, 2010), foi criado, em janeiro de 2007, para viabilizar a ação coordenada do estado em áreas de resultado definidas no PMDI, agrupando os projetos estruturadores segundo a lógica dos principais resultados esperados para a sociedade. De acordo com Busjeet (2010), havia um pressuposto de que, complementarmente à estrutura da SEPLAG, necessitava-se de um novo tipo de instituição dinâmica, reduzida e temporária, e com atuação em vários setores do governo, guiada pela compreensão geral do PMDI.

O EpR constituiu, assim, uma estrutura temporária, coordenada pelo vice-governador do estado. Além dessa estrutura formal criada legalmente, o EpR é subdividido em três unidades distintas: Unidade de Gestão Estratégica, Unidade de Empreendedores e Unidade de Indicadores (BECHELAINE, 2010). A Unidade de Gestão Estratégica responsabiliza-se pela preparação dos comitês de resultados[2] e pelo monitoramento dos planos e ação, derivados das reuniões; a Unidade de Empreendedores dedica-se ao recrutamento e acompanhamento dos empreendedores públicos[3], e, por fim, a Unidade de Indicadores é de responsabilidade da coordenação dos indicadores do PMI.

Coube ao EpR, por meio da Unidade de Indicadores, estruturar uma gestão por resultados no governo de Minas Gerais. Esse trabalho consistiu no acompanhamento dos resultados finalísticos, elencados no PMDI e no desenvolvimento de frentes de pesquisa e avaliação, dotando o governo de informações acerca de seus resultados. A Unidade de Indicadores abrangeu a tarefa de ter uma melhor memória e registro desse conjunto de indicadores constantes no PMDI em 2007. Essa tarefa consistiu na revisão do conjunto de indicadores que compõem o PMDI, tornando-os passíveis de acompanhamento anual e robusto. Além da atividade ligada à melhoria dos indicadores, coube ao EpR, por meio da Unidade de

---

[2] Tais comitês consistem em reuniões presididas pelo vice-governador, realizadas por área de resultados, abordando principalmente os projetos estruturadores daquela área, e, portanto, envolvendo os responsáveis por tais projetos (gerentes e alta administração). O objetivo dessas reuniões é gerar um acompanhamento sistematizado do vice-governador às ações do governo, possibilitando, para cada reunião, a geração de um plano de ação operativo/estratégico, com foco na implementação das políticas. Para maiores informações, ver Bechelaine (2010).

[3] Sobre o assunto empreendedores públicos ver Bernardi e Guimarães (2008).

Indicadores, desenvolver um conjunto de avaliações e pesquisas acerca de suas políticas públicas. A ideia de se criar um Sistema de Monitoramento e Avaliação (M&A) para o Governo de Minas Gerais começa a ser discutida paralelamente ao esforço de melhoria dos indicadores no EpR no início de 2007 (GUIMARÃES; CAMPOS, 2009).

Nota-se, portanto, um maior direcionamento das práticas de monitoramento das políticas públicas no período 2003-2006, principalmente devido à presença do escritório central do governo, o GERAIS. Posteriormente, no período de 2007, nota-se um direcionamento mais claro das avaliações de políticas públicas, seja pelo estabelecimento de indicadores finalísticos para o acompanhamento das mesmas, seja pelo conjunto de avaliações dos projetos estruturadores incentivado, principalmente, pelo EpR. No entanto, percebe-se que nenhum dos atores retrataram, até o momento, o que ocorreu posteriormente à conclusão dessas avaliações, não existindo trabalhos que refletem os mecanismos construídos para a utilização dos resultados do conjunto implementado.

Após analisados o papel e o desenvolvimento das avaliações de políticas públicas no Brasil e o contexto inicial em que foram introduzidas as avaliações de impacto em Minas Gerais, o capítulo seguinte irá abordar a metodologia de estudo utilizada para o desenvolvimento da etapa empírica da dissertação.

5 METODOLOGIA

Esta pesquisa busca analisar, de forma conceitual e empírica, o tema da utilização dos resultados de avaliações de políticas públicas realizadas, em especial, no governo do estado de Minas Gerais. Visa, portanto, a começar pela pesquisa bibliográfica, revisar o tema da utilização das avaliações e, a partir de pesquisa documental e empírica, identificar os propósitos, abordagens, processos e, principalmente, quais os fins das avaliações realizadas no contexto mineiro.

Para melhor responder ao problema de pesquisa, optou-se pelo método exploratório-descritivo, de acordo com o modelo proposto por Vergara (2006). Exploratório, pois visa explorar um assunto geral, sobre qual existem poucos conhecimentos no contexto nacional, e entender a dinâmica da utilização dos resultados das avaliações. Descritivo, na medida em que buscará expor os conceitos fundamentais das avaliações de políticas públicas, a influência dos fatores que interferem na utilização de seus resultados e, ainda, os processos de implantação das avaliações no governo do estado de Minas Gerais, bem como, seus principais apontamentos.

Como técnica fundamental da pesquisa propôs-se utilizar um estudo de casos múltiplos. Embora os estudos de casos múltiplos, também conhecidos como estudos comparativos, apresentem a mesma estrutura lógica dos estudos de casos únicos, as evidências resultantes de casos múltiplos são consideradas mais convincentes, e o estudo global é visto como mais robusto (YIN, 2003). Com o intuito de aumentar a confiabilidade da pesquisa, apresenta-se, a seguir, uma visão geral da técnica do estudo de casos múltiplos, das estratégias utilizadas nesta pesquisa e seus procedimentos de campo, que servirão como guia para a confecção da dissertação.

5.1 Os estudos de casos múltiplos

A necessidade dos estudos de caso surge do desejo de entender os fenômenos sociais complexos, como afirmado por Yin (2003). Os estudos de caso são utilizados em muitas situações para contribuir para o conhecimento dos fenômenos individuais, grupais, organizacionais, sociais, políticos e relacionados. Nesse sentido, o método permite que os investigadores retenham as características holísticas e significativas dos eventos da vida real –

como os ciclos individuais da vida, o comportamento dos pequenos grupos, os processos organizacionais e administrativos, a mudança de vizinhança, o desempenho escolar, as relações internacionais e a maturação das indústrias.

De acordo com Yin (2003), as principais características dos estudos de caso são duas. Primeiro, o estudo de caso constitui uma investigação empírica que analisa um fenômeno contemporâneo em profundidade e em seu contexto de vida real, especialmente quando os limites entre o fenômeno e o contexto não são claramente evidentes. Em segundo lugar, em decorrência da indefinição do fenômeno e o contexto, a pesquisa pode apresentar mais variáveis de interesse do que pontos de dados e, como resultado, contar com múltiplas fontes de evidência. Para o autor, em geral, os estudos de caso são o método preferido quando: 1) os problemas de pesquisa têm questões *como* ou *por que*; 2) o investigador tem pouco controle sobre os eventos; 3) o enfoque ocorre sobre um fenômeno contemporâneo no contexto da vida real.

Os estudos de caso têm várias aplicações na pesquisa. Uma opção é explicar os presumidos vínculos causais nas intervenções da vida real que são demasiadamente complexos para as estratégias de levantamento ou experimentais. Uma segunda aplicação é descrever uma intervenção e o contexto da vida real no qual ela ocorreu. Em terceiro lugar, os estudos de caso podem ilustrar determinados tópicos em uma avaliação, novamente em um modo descritivo. Como quarta e última aplicação, a estratégia de estudo de caso pode ser usada para explorar as situações em que a intervenção, sendo avaliada, não aponta um único e claro conjunto de resultados.

Como a utilização dos resultados das avaliações, objeto deste projeto, constitui um fenômeno ainda pouco estudado no Brasil, como abordado na justificativa desta dissertação, a aplicação da técnica de estudo de caso torna-se, assim, propícia e adequada às circunstancias da pesquisa.

A pesquisa do estudo de caso inclui, tanto estudos de caso único, quanto de casos múltiplos, também conhecidos como estudos comparativos, que são, na realidade, apenas duas variantes dos projetos de estudos de caso (YIN, 2003). A justificativa para a escolha de estudo de casos múltiplos se deu, principalmente, pelas vantagens apresentadas e pela utilidade da técnica como forma de investigação. Para Yin (2003), nos estudos de casos múltiplos, as evidências

são consideradas mais convincentes e o estudo global é visto como mais robusto, uma vez que observam uma lógica de replicação em diferentes contextos analisados. De acordo com o autor, embora todos os projetos possam levar a estudos de caso único bem-sucedidos, quando houver escolha (e recursos), é preferível optar por estudo de casos múltiplos em vez de projetos de caso único.

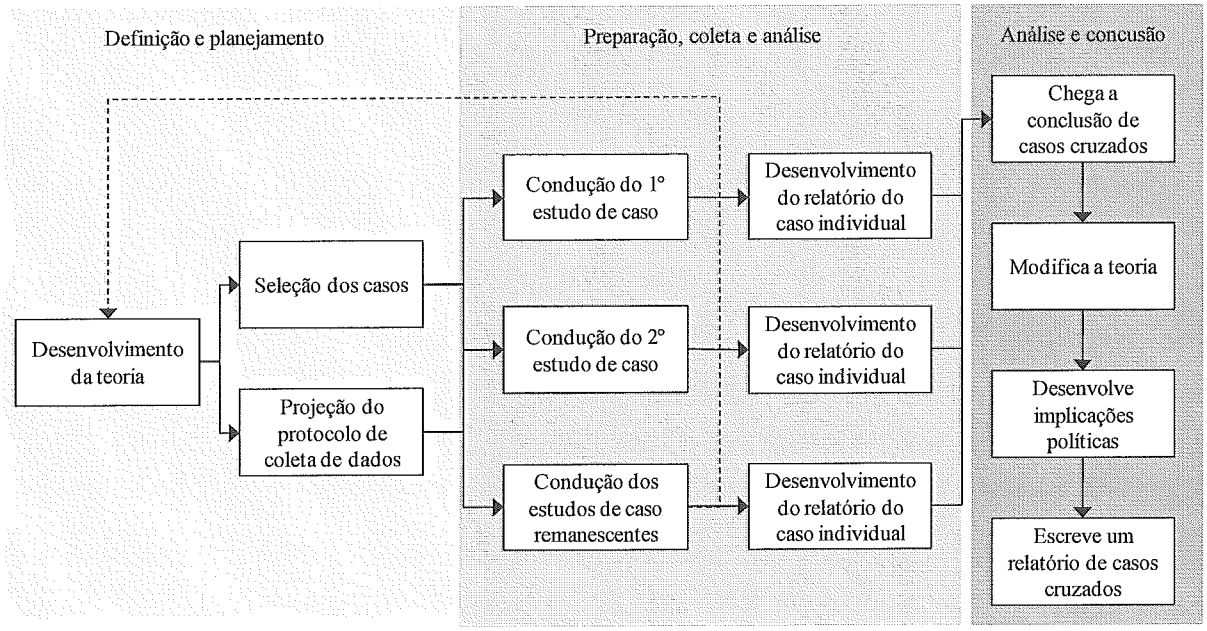
Um passo importante em todos os procedimentos de replicação é o desenvolvimento de uma rica estrutura teórica, na qual são expostas as condições em que seja possível encontrar um fenômeno em particular (uma replicação literal), assim como as condições em que não seja provável que se encontre (uma replicação teórica). Essa lógica de replicação, se aplicada a experimentos ou a estudos de caso, deve ser diferenciada da lógica de amostragem comumente utilizada em levantamentos de dados. Qualquer aplicação da lógica de amostragem aos estudos de caso estaria mal direcionada, pois: 1) os estudos de caso, em geral, não devem ser usados para avaliar a incidência de fenômenos; 2) um estudo de caso teria que tratar, tanto do fenômeno de interesse, quanto de seu contexto (grande número de variáveis, grande número de casos); 3) se uma lógica de amostragem tivesse de ser aplicada a todos os tipos de pesquisa, muitos tópicos não poderiam ser empiricamente investigados.

Torna-se importante mencionar, contudo, que, como apontado por Yin (2003), o fato de que um determinado projeto exige estudos de casos múltiplos não elimina a variação identificada nos casos únicos: cada caso em particular ainda pode ser holístico ou incorporado (casos múltiplos holísticos ou casos múltiplos incorporados). O autor explica que o caso único pode ser um programa que envolve um grande número de projetos financiados que seriam, então, as unidades incorporadas, selecionadas por técnicas de amostragem ou agrupamentos. O projeto resultante é chamado de projeto de estudo de caso incorporado. No caso oposto, se o estudo de caso examina apenas a natureza global de um programa ou de uma organização, o projeto se caracterizaria como holístico. Este tipo de estudo torna-se vantajoso quando não é possível identificar nenhuma subunidade lógica ou mesmo quando a própria teoria subjacente ao estudo é holística. No entanto, essa abordagem global pode fazer com que o pesquisador não examine com maiores detalhes o fenômeno, sendo o caso conduzido de maneira abstrata. A diferença entre esses dois projetos depende do tipo de fenômeno que será estudado e das questões de pesquisa. Cada caso deve servir a um propósito específico dentro do escopo global da investigação, afirma Yin (2003).

Neste trabalho, considerando as diferentes avaliações de projetos realizadas pelo governo de Minas Gerais, conforme abordado por Ferreira, Lopes e Campos (2010), optou-se pela realização de estudos de casos múltiplos holísticos. Tal escolha deve-se à existência de características e contextos diferenciados em cada uma das avaliações de políticas realizadas pelo governo de Minas Gerais. Como será observado na seção seguinte, que aborda a seleção dos casos, cada uma das avaliações tem justificativa, contexto, atores e lógica de organização próprias, perfazendo casos únicos e diferenciados, mas que abordam o mesmo questionamento, qual seja, o da utilidade das avaliações para o aprimoramento dos projetos.

A técnica de análise por estudo de casos múltiplos deve ser conduzida em três etapas: 1) definição e planejamento; 2) preparação, coleta e análise; 3) análise e conclusão (YIN, 2003). O primeiro passo consiste no desenvolvimento da teoria, seleção dos casos e unidades de análise para a condução posterior de cada caso individual. Somente após a confecção de pesquisas individuais dos casos é possível verificar a lógica de replicação existente entre eles mesmos e a teoria, como pode ser observado na figura 2.

Figura 2 – Método de estudos de caso múltiplos



Fonte – Yin, 2003, adaptado pela autora da dissertação.

Assim, ao seguir os passos de um estudo de casos múltiplos como descrito por Yin (2003), o presente trabalho será desenvolvido em três etapas principais. A primeira delas compreende o desenvolvimento desta dissertação que expõe a definição das questões da pesquisa e seu

embasamento teórico, assim como a demarcação dos casos selecionados e dos procedimentos de coleta e análise de dados a serem utilizados na pesquisa. Após a aprovação desta dissertação, iniciar-se-á a etapa de preparação, coleta e análise de dados, na qual cada caso selecionado será conduzido como se fosse um único caso específico. Por fim, a última etapa da pesquisa da dissertação consistirá em observar os pontos comuns e díspares dos casos específicos analisados a fim de encontrar implicações para a teoria de utilização dos resultados de avaliações de políticas públicas. A seguir, definem-se os casos que comporão o estudo de casos múltiplos.

5.2 Seleção dos casos

De acordo com Yin (2003), permanece uma dificuldade fundamental para se definir o que é um *caso*, uma vez, que não existe uma única possibilidade. Como orientação geral, Yin (2003) argumenta que a tentativa de definição do caso deve estar relacionada à maneira como foram definidas as questões iniciais da pesquisa.

Para melhor responder o problema de pesquisa *como os diferentes stakeholders envolvidos nos processos de avaliação de políticas públicas realizadas pelo governo do estado de Minas Gerais percebem a utilidade dos resultados das avaliações? Por quê?* optou-se pela escolha de casos das avaliações de projetos estruturadores realizadas pelas diferentes secretarias de estado que compõe a estrutura governamental de Minas Gerais. Dessa forma, os casos a serem selecionados dizem respeito exclusivamente às avaliações de impacto dos projetos estruturadores implementadas no período de 2007 a 2010.

A delimitação de tempo escolhida deve-se ao fato que, de acordo com as pesquisas bibliográficas, somente a partir do ano de 2007 foi possível observar um claro direcionamento do governo mineiro para contratação e desenvolvimento na área de avaliações (FERREIRA; LOPES; CAMPOS, 2010). Embora existissem iniciativas anteriores, como a avaliação anual do PPAG, assim como iniciativas isoladas de alguns programas, foi a partir de 2007 que se notou maior interesse pelas avaliações, principalmente por aquelas relacionadas à efetividade (avaliações de impacto) dos programas. Em outras palavras, somente a partir desse período, foi possível encontrar evidências da preocupação de governantes em contratar avaliações.

A unidade de análise escolhida corresponde, portanto, às "avaliações de projetos estruturadores realizadas pelo Governo do Estado de Minas Gerais, a partir do ano de 2007". No Quadro 7 apresenta-se a relação das avaliações dos projetos estruturadores (possíveis casos) realizadas a partir do ano de 2007, de acordo com Ferreira, Lopes e Campos (2010). Tais avaliações dizem respeito a projetos de diferentes áreas de atuação, inseridas em contextos organizacionais também diferenciados, uma vez que foram realizadas por Secretarias de Estado diversas e conduzidas por avaliadores de centros de pesquisas distintos.

Quadro 11 - Relação dos programas avaliados

Secretaria de Estado	Projeto avaliado	Instituição responsável pela avaliação	Tipo de avaliação realizada
Secretaria de Estado da Educação (SEE)	Programa de Educação Profissional	CEDEPLAR	Avaliação de Impacto
	Escola em Tempo Integral	CEDEPLAR	Avaliação de Impacto
	Aceleração da Aprendizagem	CEDEPLAR	Avaliação de Impacto
Secretaria de Estado de Saúde (SES)	Viva Vida	FJP	Avaliação de impacto e processos
	Pró-Hosp	FJP	Avaliação de impacto e processos
Secretaria de Estado de Defesa Social (SEDS)	Gestão Integrada de Ações e Informações do Sistema de Defesa Social	FJP	Avaliação de Impacto
Secretaria de Estado de Agricultura (SEAPA)	Minas Sem Fome[1]	FJP	Avaliação de Impacto
Secretaria de Estado de Desenvolvimento Social (SEDESE)	Poupança Jovem	CEDEPLAR	Avaliação de Impacto

Fonte – governo do estado de Minas Gerais, 2007, adaptado pela autora as dissertação.

Notas: [1] Projeto executado pela Empresa de Assistência Técnica e Extensão Rural do Estado de Minas Gerais (EMATER), vinculada à SEAPA.

Ao considerar o tempo disponível para a realização desta pesquisa, que infelizmente não permite aprofundar o estudo de todas as avaliações identificadas, e a complexidade do tema, optou-se por selecionar, inicialmente, quatro avaliações de projetos realizadas por diferentes Secretarias de Estado e também distintas instituições contratadas para realizar as avaliações. As avaliações selecionadas para compor este estudo de casos múltiplos constituem:

- Avaliação do projeto estruturador Aceleração da Aprendizagem, da SEE;
- Avaliação do projeto estruturador Viva Vida, da SES;
- Avaliação do projeto Gestão Integrada de Ações e Informações do Sistema de Defesa Social, da SEDS;
- Avaliação do projeto Minas Sem Fome, da SEAPA, executado pela EMATER.

A justificativa da escolha dos casos deve-se à análise de fatores combinados. Em primeiro lugar, observa-se, por exemplo, que SEE e SES foram as secretarias de Estado que consumiram maior volume de recursos com avaliações de programas no estado. É importante ressaltar também que foram as primeiras secretarias a contratar avaliações e, assim, contaram com os resultados das mesmas para aprimorar seus programas. A escolha das demais avaliações realizadas pelas SEAPA e SEDS deve-se à necessidade de observar a inserção de avaliações em contextos organizacionais diferenciados e variados. Como o estudo aborda a utilização de tais resultados, torna-se essencial averiguar um espaço de tempo entre a conclusão das avaliações e a realização desta pesquisa – o que pode ser verificado em todos os casos. Por fim, atenta-se para características como disponibilidade de informações e a facilidade de acesso às mesmas, favorecendo a solicitação dos dados e informações necessárias ao desenvolvimento desta pesquisa.

Dessa forma, os casos selecionados correspondem a avaliações de projetos estruturadores Aceleração da Aprendizagem, Viva Vida, Gestão Integrada de Ações e Informações do Sistema de Defesa Social e Minas Sem Fome realizadas pelas secretarias de estado de Educação, Saúde, Defesa Social e Agricultura, respectivamente, a partir do ano de 2007 no governo do estado de Minas Gerais.

No entanto, cabe ressaltar que, embora se tenha optado por uma unidade de análise, a mesma não será considerada definitiva. Como afirmado por Yin (2003), a escolha da unidade de análise pode ser revisitada a partir de descobertas que surgiram durante a coleta dos dados. Nesse sentido, uma vez que tenha sido estabelecida a definição geral do caso, torna-se importante fazer novas considerações da unidade de análise.

5.3 Coleta e análise de dados

A preparação para a coleta dados de deve ser feita de forma correta para que não se coloque em risco todo o trabalho de investigação do estudo de caso (YIN, 2003). Como apontado pelo autor, os procedimentos necessários para a coleta de dados em um caso único convergem para o protocolo dos casos múltiplos, não havendo diferenças significativas. Yin (2003) deixa claro que um bom estudo de caso utiliza o maior número possível de fontes, sendo que nenhuma fonte tem vantagem sobre as outras, são apenas complementares.

Tendo em vista que a natureza do objeto de pesquisa demanda uma análise de caráter interpretativo dos dados coletados, pretende-se utilizar estratégia qualitativa. As evidências deste estudo correspondem a fontes distintas, quais sejam: 1) observação direta; 2) análise documental; 3) entrevistas.

A observação direta consiste em realizar visitas de campo, principalmente, às secretarias de Estado que participaram dos casos de avaliações selecionados e às instituições contratadas para aplicar as avaliações. De acordo com Yin (2003), com a observação direta, cria-se oportunidade para fornecer informações adicionais sobre o tópico que está sendo estudado.

O uso mais importante dos documentos é corroborar e valorizar as evidências oriundas de outras fontes e fornecer outros detalhes específicos (YIN, 2003). Devido ao seu valor global, os documentos desempenham um papel explícito em qualquer coleta de dados, ao se realizar estudo de caso. Assim, tendo como unidade de análise as avaliações de políticas de alguns projetos estruturadores selecionados, serão consultados documentos como os contratos e termos de referência elaborados para a contratação das instituições responsáveis pela avaliação, projetos de avaliação (caso existam), relatórios finais e intermediários das avaliações, assim como demais documentos relacionados ao tema e ao processo de avaliação.

Além da observação direta e da análise documental, constitui uma das mais importantes fontes de informação para este estudo de casos múltiplos as entrevistas realizadas, em especial, com os potenciais usuários dos estudos de avaliação. Uma vez que se procura construir a percepção de diferentes atores quanto à utilidade das avaliações desenvolvidas no âmbito do governo de Minas Gerais, as entrevistas mostram-se como processo essencial ao alcance dos objetivos deste trabalho. Uma etapa inerente à realização das entrevistas consiste em definir quem são os atores entrevistados na pesquisa. No entanto, apresenta-se o seguinte questionamento: quem são os *stakeholders* envolvidos no processo de utilização de informações das avaliações e sua implementação?

De acordo com Weiss (1998a), no início da *era* das avaliações, os avaliadores acreditavam que os principais utilizadores dos estudos seriam os patrocinadores dos programas (aqueles responsáveis pelo seu orçamento), os coordenadores dos programas (aqueles que administram os programas) e, talvez, os próprios técnicos que constituem as equipes dos programas (também chamados de *staff*, em contato direto com os beneficiários). No entanto, a partir do

momento em que o campo avaliações de programas foi se desenvolvendo, além desses usuários imediatos, os avaliadores perceberam uma ampla multiplicidade de outros grupos de atores que também poderiam se beneficiar dos resultados das avaliações (WEISS, 1998a). Dentre os novos atores estão os coordenadores de programas similares que deveriam apreender como aprimorar os seus programas com as avaliações de outros, os formuladores de políticas no legislativo, bem como cientistas sociais. Weiss (1998a) também inclui os próprios avaliadores como os principais responsáveis e interessados na utilização dos resultados das avaliações e a sociedade civil organizada. É claro que nem todos os estudos avaliatórios irão atrair e interessar a todas as possibilidades de usuários das avaliações, ou irão merecer a sua atenção.

Dessa forma, como parte do estudo empírico e baseando-se no entendimento da Weiss (1998a), efetuaram-se entrevistas semiestruturadas com os usuários imediatos das avaliações realizadas pelas secretarias de Estado selecionadas. Em conjunto, foram realizadas vinte e cinco entrevistas, sendo: quatro entrevistas com os gerentes dos projetos estruturadores do estado de Minas Gerais submetidos às avaliações; cinco entrevistas com as equipes dos programas envolvidas com as avaliações; dez entrevistas com avaliadores externos da Fundação João Pinheiro (FJP) e do Centro de Desenvolvimento e Planejamento Regional de Minas Gerais (CEDEPLAR) responsáveis por realizar as avaliações; três entrevistas envolvendo *sponsors* das avaliações, ou seja, pessoas envolvidas nos processos de financiamento e contratação, como os coordenadores do EpR e da FJP e, por fim, mais três entrevistas com atores relacionados aos projetos de maneira indireta, como exemplo dos técnicos do EpR que acompanharam o desenvolvimento do conjunto de avaliações como se vê no quadro 8.

Quadro 12 – Relação dos entrevistados

Entrevistados	Descrição	Número de entrevistados
Coordenadores/gerentes de programas	Constituem atores estratégicos do projeto, detentores de informações sobre todos os aspectos dos mesmos. Geralmente participam da contratação dos processos avaliatórios e são os mais interessados nos resultados	4
Equipe dos projetos	Constituem atores de nível tático do projeto, são responsáveis por assessorar os gerentes dos projetos e acumulam diferentes fincões no projeto	5
Avaliadores externos	São os atores que mais conhecem o processo de avaliação, bem como o conteúdo dos relatórios. Nos estudos de caso escolhidos, podem ser representados por membros da FJP e CEDEPLAR – instituições contratadas para realizar as avaliações	10
Sponsors	Constituem diferentes tipos de atores. Podem ser políticos, formuladores de políticas, entre outros atores financiadores das avaliações e/ou envolvidos nos processos de contratação. No caso de Minas Gerais, constituem os organismos internacionais, como o BIRD e BID, e os patrocinadores internos, como os coordenadores do EpR e da FJP	3
Responsáveis nos órgãos de planejamento do governo	Constituem atores relacionados aos projetos de maneira indireta, ou seja, apesar de conhecer os projetos e monitorá-los periodicamente, não participam da implementação das ações. Exemplos deste tipo de ator no governo de Minas Gerais podem ser encontrados nos setores que cuidam do orçamento, Acordo de Resultados, Secretaria de Planejamento, Estado para Resultados	3

Fonte – Elaborado pela autora da dissertação.

É importante ressaltar que todos os entrevistados assinaram um Termo de Consentimento Livre e Esclarecido em função de sua participação voluntária na pesquisa. Nessa oportunidade esclarece-se que as informações seriam utilizadas somente para os fins desta pesquisa e tratadas com o mais absoluto sigilo e confidencialidade, de modo a preservar a sua identidade. Os nomes dos entrevistados foram, assim, substituídos por indicativos de como se deu a sua participação na pesquisa, a fim de identificar a natureza de participação (avaliador, gestor etc) em relação à avaliação.

Quanto às dimensões da análise das entrevistas realizadas, pretende-se reproduzir questões semelhantes para os diferentes atores envolvidos nos casos de avaliações, com a finalidade de comparar diferentes visões sob a ótica das avaliações feitas. Tais dimensões foram desenvolvidas de acordo com a seção 3.3 desta dissertação, cujo objetivo foi levantar, na visão de diferentes autores, os fatores e variáveis que influenciavam a utilização das avaliações como, por exemplo, a relevância do contexto em que a pesquisa avaliatória está inserida, a interação entre os vários atores envolvidos na avaliação, a qualidade dos relatórios

etc. Dessa forma, por meio do roteiro de entrevistas, que se encontra no apêndice deste trabalho, procura-se identificar quais os atores relevantes na gestão das avaliações e verificar como orientam seus cursos de ação com relação aos processos e sistemas de avaliação: suas preferências, isto é, ideias e interesses de que são portadores esses agentes; grau de acesso à informação; uso da avaliação e da informação disponibilizada para a inserção de mudanças. O questionário abarcou questões ligadas ao funcionamento do monitoramento e avaliação no governo, a forma pela qual o trabalho da equipe contribui nesse processo, principais características e instrumentos utilizados, relacionamento com outras equipes, além dos benefícios e dificuldades identificadas. Formadas essencialmente por perguntas abertas, obtiveram-se, aproximadamente, quarenta e uma horas de gravação de entrevistas, com uma média de tempo, portanto, de uma hora e quarenta e dois minutos por entrevistado, de acordo com registros próprios.

Em relação aos dados qualitativos coletados, utilizou-se o método de análise de conteúdo que, de acordo com Bardin (1988) consiste na organização das informações em torno de três polos cronológicos: 1) pré-análise; 2) exploração do material; 3) tratamento dos resultados, inferência e interpretação. Cada uma dessas etapas tem uma série de procedimentos que devem ser adotados visando tirar das entrevistas e materiais coletados as informações que responderão ao problema de pesquisa. Segundo Bardin (1988, p.101), os resultados brutos obtidos "[...] são tratados de maneira a serem significativos (falantes) e válidos. [Desse modo, serão utilizadas operações estatísticas como percentagens e análises fatoriais que] permitem estabelecer quadros de resultados, diagramas, figuras e modelos, os quais condensam e põem em relevo as informações fornecidas pela análise".

5.4 Dificuldades e limitações do estudo

As limitações do estudo estão relacionadas tanto ao tema estudado, que diz respeito à utilização das avaliações realizadas, como ao método de pesquisa selecionado. Em primeiro lugar, tratando-se das dificuldades e entraves inerentes ao estudo da utilização das avaliações, Leviton e Hughes (1981) ressaltam três pontos principais, já comentados anteriormente neste trabalho: 1) ausência de documentação sobre os processos de utilização após a conclusão de uma avaliação, uma vez que as avaliações são utilizadas de maneira informal; 2) dificuldade de provar que as mudanças implementadas por um programa não seriam implementadas caso a avaliação não fosse realizada, ou seja, na ausência da informação da avaliação, os

envolvidos nas atividades do programa avaliado teriam pensado ou agido de forma diferente; 3) indefinição quanto à unidade de análise *utilização*, uma vez que não se sabe ao certo qual o melhor exemplo de utilização e que não existem medidas adequadas para apurar uma *taxa de uso*.

Quanto às limitações próprias ao método de pesquisa, atenta-se, primeiramente, para a natureza restrita deste trabalho, já que o estudo foi realizado apenas para algumas avaliações de impacto de programas governamentais e, portanto, não pode ser generalizado. Ainda que alguns fatores observados na pesquisa se reproduzam em mais de um estudo de caso, observa-se que o universo escolhido não pode representar o ocorrido em demais órgãos da administração pública mineira ou brasileira.

As dificuldades operacionais do estudo estão relacionadas à dificuldade de encontrar as pessoas envolvidas com as avaliações na época em que aconteceram devido a dois motivos distintos. Primeiro, por causa do número reduzido de pessoas das equipes dos projetos avaliados que participaram e tiveram conhecimento da realização das avaliações. Embora fosse fácil encontrar os nomes dos avaliadores envolvidos nas avaliações, verificou-se certa dificuldade em encontrar pessoas ligadas à avaliação nas secretarias do governo, já que alguns gestores dos projetos avaliados confessaram que não poderiam indicar mais pessoas da equipe, pois elas não se envolveram com os estudos avaliatórios.

Um segundo ponto está relacionado à dificuldade de encontrar os entrevistados selecionados, dado que muitos deles não trabalhavam mais nas mesmas funções de quando as avaliações foram desenvolvidas. Uma vez que os ciclos de avaliações de impacto geralmente são longos, e decorreu um tempo de quatro anos desde o início das avaliações até a realização desta pesquisa, muitos dos entrevistados não se encontravam nos mesmos postos de trabalho. Além disso, é possível verificar quebras de participação, sendo que um entrevistado participou da contratação da avaliação, saiu do posto, e outro entrevistado participou de outra fase do estudo como, por exemplo, a entrega dos resultados. Observa-se, portanto, uma falta de memória de uma experiência relativamente nova devido à rotatividade dos atores envolvidos.

Por fim, cabe a ressalva de que a parcialidade do pesquisador, que participou como coadjuvante do processo, pode, em alguma medida, influenciar o resultado da análise,

considerando o seu conhecimento e experiência profissional no assunto de avaliações de políticas. A seguir, apresentam-se os resultados da análise desta pesquisa.

6 AS AVALIAÇÕES DOS PROJETOS ESTRUTURADORES NO GOVERNO DE MINAS GERAIS

De acordo com Busjeet (2010), a introdução da orientação por resultados em um governo pode ser estimulada por diferentes fatores. E as avaliações de políticas públicas, que constituem parte deste modelo de orientação por resultados, também não ocorrem ao acaso. Fatores contextuais específicos, aspectos políticos e institucionais, assim como momentos de lideranças únicas, são alguns dos aspectos para o desenvolvimento e evolução de sistemas de Monitoramento e Avaliação.

No caso de Minas Gerais, a introdução de um sistema de monitoramento e avaliação é fruto de uma convergência de fatores que se destacaram a partir de 2003. Dentre os fatores determinantes para a assunção das avaliações de políticas, enfatiza-se a disciplina fiscal – também conhecida como Choque de Gestão, a criação de um escritório central para o monitoramento intensivo de projetos, o processo de contratualização de resultados e a mensuração de indicadores finalísticos. Neste capítulo apresentam-se a trajetória das avaliações realizadas no período de 2007 a 2010 já construídas a partir da visão dos entrevistados, como um marco introdutório antes de entrar nos capítulos específicos de cada uma das avaliações realizadas, afirmados nos capítulos sete a dez.

6.1 Bases do movimento de avaliação e atores envolvidos

Embora muitas ferramentas houvessem sido criadas, como o monitoramento intensivo de projetos, assim como a contratualização de resultados como os acordos de resultados e o acompanhamento e apuração de indicadores de desempenho, Ferreira, Lopes e Campos (2010) afirmam que uma abordagem sistemática de avaliações de políticas públicas ainda é algo recente no estado de Minas Gerais. Muitos dos estudos avaliatórios que já ocorreram nos governos mineiros anteriores constituíram iniciativas isoladas e, muitas vezes, influenciada por organismos internacionais de desenvolvimento como o Banco Mundial (BIRD) e o Banco interamericano (BID), como relata o entrevistado do EpR. Na percepção do entrevistado da Secretaria de Estado de Educação (SEE) e do *staff* do EpR, a execução de um conjunto de avaliações de impacto foi uma iniciativa inovadora e, dada a sua natureza pioneira, seria também natural que o processo apresentasse dificuldades.

> O governo não tem maturidade e não tem histórico de avaliação. A avaliação sempre aparece em termos de disciplina e de conteúdo de maneira esporádicas, esforços individuais ou bem pontuais ao longo do tempo de maneira diferente. Então você pega para trás e você vê que tem várias tentativas de avaliação de alguns programas e tal e muito fomentado pelos órgãos multilaterais, Banco Mundial e BID principalmente, e que isso é uma coisa que nunca teve uma cultura especifica no Governo. (COORDENAÇÃO EPR1, grifo da autora da dissertação)[4]

> Eu acho que foi uma iniciativa até um pouco inovadora. A gente tentou até pesquisar e não tem nenhum registro aqui na secretaria de avaliação de impacto ou alguma política pública feita naquele modelo. Eu não conheço todas as avaliações de impacto, mas naquele modelo nunca tinha sido feito não. Foi a primeira, e sempre a primeira é um pouco mais difícil. (EQUIPE SEE2)

> De positivo eu acho que foi bastante ousado fazer as avaliações na época mesmo sabendo que provavelmente haveria problemas como cumprimento de prazo, como obtenção de dados necessários, como envolvimento dos atores; e ainda sim optou-se por fazer algo que não era feito antes: avaliações de impacto com as metodologias recomendadas pela literatura, com participação de avaliadores externos e participação de agentes internos ao Governo que fossem intermediários. Então, eu acho que foi muito válido o esforço nessa tentativa de amadurecer o sistema de monitoramento e avaliação de Minas Gerais. (TÉCNICO EPR3, grifo da autora da dissertação)

Assim, mesmo que alguns estudos avaliatórios de programas específicos e em períodos particulares da história de governos mineiros já tivessem sido realizadas, o surgimento de uma agenda de avaliações sistemática inicia-se na segunda geração do choque de gestão (no caso em questão, 2007-2010), ocasião em que se observou uma orientação por resultados mais forte na gestão pública. Verifica-se, contudo, que o caminho que caracterizou a experiência mineira no que tange à implementação das avaliações de políticas públicas foi um processo construído de forma gradual, *construído à medida que foi se apresentando*. De acordo com o entrevistado EpR1, um conjunto de circunstâncias únicas, fatores excepcionais combinados, como o fomento do Banco Mundial, a liderança da FJP e a criação da capacidade de avaliar culminaram na execução do conjunto de avaliações investido a partir de 2007.

> Obviamente isso não foi um movimento 100% planejado no sentido de vamos sair daqui e chegar lá, a gente foi construindo um caminho na medida em que ele foi se apresentando. Por quê? Porque a gente teve o Banco Mundial, que era fomentador disso, a gente teve um momento único de liderança da Fundação João Pinheiro, de ter uma pessoa que reconhece essa importância – tem muito a ver com a liderança das pessoas também isso – e você teve o momento de criar capacidade de avaliar. Porque não basta querer avaliar, assim; a partir de hoje tudo vai ser avaliado. (COORDENAÇÃO EPR1)

---

[4] Nesta e nas demais transcrições de entrevistas não se levará em conta a correção linguística e gramatical. O objetivo é preservar a espontaneidade das falas dos entrevistados.

Assim, observa-se que a conjuntura para o desenvolvimento das avaliações teve sua origem em razão de mudanças introduzidas em períodos anteriores a 2007, até que a variável resultados passasse a desempenhar um importante papel no governo.

De acordo com Busjeet (2010), a articulação dos diferentes papéis e responsabilidades do modelo de gestão por resultados mineiro foi conduzida por três entes estaduais distintos, com inúmeros *stakeholders* envolvidos e, em consequência, interesses também diferentes. O primeiro destes é a Secretaria de Estado de Planejamento e Gestão (SEPLAG), onde funcionam as equipes do SCGERAES – responsável pelo monitoramento dos projetos estruturadores, da SUMIN – responsável pela coordenação e gestão dos acordos de resultados, e da SCPPO – responsável por incorporar as informações sobre resultados nos processos orçamentários. Além da SEPLAG, integram o conjunto o EpR, a FJP e o Banco Mundial, como instituição externa fomentadora do processo. Um dos avaliadores do projeto avaliado Minas Sem Fome (MSF) comenta quais os atores participaram da iniciativa.

> Participaram da iniciativa a Fundação João Pinheiro, que estava criando um núcleo de avaliação; a SEPLAG, por meio desse Estado para Resultados; e tinha o Banco Mundial, porque claramente isso estava dentro de um contexto do Choque de Gestão e que se pensava em um novo patamar de política pública, de gasto público que não seria só assentado na entrega, mas também na efetividade, então teria de incorporar a avaliação de políticas públicas. Então, claramente havia uma indução do Banco Mundial e ele foi muito parceiro dentro desse processo. (AVALIADOR MSF4)

A introdução da agenda de avaliações do governo mineiro que, na visão dos entrevistados, constituía um movimento pioneiro de realizar avaliações das políticas de forma sistemática, nasceu de um conjunto de circunstâncias únicas. Para os entrevistados do EpR e do CEDEPLAR, o momento se deu pela soma de lideranças do governo e da FJP, pela presença do Banco Mundial e pelo diálogo que fluiu.

> Juntou a liderança do Governo, o espaço do Banco Mundial e a especialidade do Banco Mundial que foi reconhecida e teve espaço, com uma capacidade da liderança da Fundação João Pinheiro que via a importância de se ter essa questão. Havia diálogo e pessoas que sabiam do que se tratava a avaliação e conseguiram dialogar sobre o assunto para fazer o assunto render. (COORDENAÇÃO EPR1)

> Havia muito mais discussão e interesse no início. Nessa época, vinham os caras do Banco Mundial. Os caras do Banco Mundial puxavam. Os americanos vinham e tal, tinha aquela discussão e a coisa ficava muito mais interessante. Na verdade, foi um período interessante porque o vice-governador tinha muito interesse nessas coisas. Você tinha um presidente na Fundação João Pinheiro que gostava disso, então você tinha todo o mundo, um monte de economistas, gente interessada no *troço*, nesse

> tipo de abordagem. Foi muito legal no início, foi muito promissor. (COORDENAÇÃO CEDEPLAR)

De acordo com Busjeet (2010), a FJP constitui uma agência estadual criada em 1969 para munir o governo do estado de estudos econômicos, sociais e lançamento e análise de estatísticas oficiais, além das funções de treinamento de servidores públicos. Para o entrevistado EpR1 e para o avaliador VV2 (um dos avaliadores do projeto Viva Vida), a FJP sempre tem em suas funções o cerne da avaliação de políticas, mas nem sempre tal movimento foi constante, o que diferencia as avaliações esporádicas já realizadas pela instituição, da prática institucional que estava sendo criada, como apontam o primeiro e segundo trecho de entrevistas selecionados. Na visão do entrevistado da coordenação da FJP, havia um direcionamento de tornar a FJP uma referência no quesito avaliações de programas.

> A Fundação é um braço que sempre trabalhou a ideia de avaliação, mas, como a questão do governo, ela teve seus momentos bons e ruins, e ora vinha o conforto, ora se perdia e você não tinha uma sistemática. Então, o que a gente introduziu na medida em que foi dada essa ênfase foi tentar criar uma maneira mais sistemática de tratar a avaliação do Governo. (COORDENAÇÃO EPR1)

> Na verdade, é importante se dizer o seguinte: a Fundação já fez outras avaliações. Fazer avaliação de políticas é uma das áreas da Fundação João Pinheiro. Eu mesma já participei de outras avaliações. O que a gente não tinha ainda feito, o que caracterizou esse período, eu diria, foi que, no mesmo período foram feitas várias avaliações e foi criado um núcleo. Então, é assim, esporadicamente surge uma avaliação. A diferença é que naquele momento, vamos dizer assim, teve interesse de criar aqui um núcleo de avaliação, talvez porque o Programa Para Resultados né, ele também estava interessado. (AVALIADOR VV2)

> Nessa mesma direção a Fundação João Pinheiro concordou em sediar, não sei se foi a segunda ou terceira rede latino-americana de políticas públicas. Mas isso foi na mesma direção de tornar a Fundação João Pinheiro um centro de referência na avaliação de políticas públicas no Brasil. O seminário da nova rede latino-americana foi sediado aqui com esse objetivo. Na reunião da rede, criou-se a Rede Brasileira de Avaliação de Políticas Públicas, que foi criada aí, a Fundação João Pinheiro teve um papel importante nisso. (COORDENAÇÃO FJP, grifo da autora da dissertação)

Para Ferreira, Lopes e Campos (2010), o esforço desenvolvido na frente de avaliações de políticas públicas foi liderado pela FJP e pelo EpR, contando com o contínuo apoio da SEPLAG e com a rede de especialistas do Banco Mundial. Para os autores, a construção do modelo de gestão por resultados mineiro foi muito apoiada pelo Banco Mundial, por meio de financiamento e, principalmente, via interação entre seus especialistas e a equipe de governo. Como afirmado pelo entrevistado EpR1, o contrato de financiamento com o Banco Mundial não foi firmado nos moldes tradicionais, ou seja, com uma contrapartida financeira do governo mineiro, mas com um incentivo grande pelo cumprimento de metas e aferição de

resultados. E parte das metas negociadas no contrato elencava a agenda de avaliações planejada pelo governo. Nas palavras do entrevistado da coordenação e dos *staff* técnico do EpR, a ideia seria colocar o cumprimento de metas como contrapartida e, dentre essas metas, a consecução de avaliações de impacto.

> O Banco Mundial, ele fez um empréstimo baseado em atingimento de metas. Ao invés de você fazer o empréstimo que você tinha que está contrapartida financeira, então, o Governo põe 100 mil, o Banco Mundial põe 100 mil. Então o Governo falou: "O Banco Mundial põe 100 mil e eu cumpro os resultados. Se eu cumprir esse quadro de metas eu recebo mais 100 mil". Então, ao elencar essas metas, parte delas foram esses processos de entregar essas avaliações. (COORDENAÇÃO EPR1)

> Não lembro o ano, acho que foi 2007 mesmo ou 2008, a gente conseguiu um empréstimo com o Banco Mundial, e a contrapartida desse empréstimo era a execução de algumas atividades, de alguns projetos. Mas isso tudo foi negociado. Acho que era interesse do Estado Para resultados fazer a avaliação de política pública: "Ah, vamos botar uma avaliação de impacto", como contrapartida. (TÉCNICO EPR1)

De acordo com o Guimarães e Campos (2009), se, por um lado, o governo havia dado bons passos no monitoramento de seus projetos estruturadores por meio do SCGERAES, pouco se sabia sobre os impactos que poderiam ser esperados pelos projetos, a magnitude desses e a existência de causalidade entre as ações e os resultados esperados. O entrevistado EpR1 e o avaliador MSF4 corroboram o pensamento dos autores, seria necessário dar uma passo a diante do simples monitoramento dos projetos.

> Bom, a gente estava na virada do Governo, entre a ideia do Choque de Gestão e a ideia de Estado para Resultados. Então a grande mudança de um modelo para o outro foi a introdução da variável resultado. Por quê? Porque por mais que o Choque de Gestão foi o marco, ele não estava muito atrelado a ideia de resultado ainda, e isso foi amadurecendo ao longo desse período culminando com esse conceito que vigorou. Então, na medida em que colocou o resultado no centro da atenção de todo mundo, várias frentes de resultados foram observadas: um forte braço disso chama-se monitoramento, que é a ideia de toda essa questão de acompanhar indicador, de criar sistemática de transparência. Tem muito a ver com o que o GERAES já tinha amadurecido nesses quatro anos e um outro pedaço disso tem a ver com a avaliação, que é saber se o que está fazendo está levando para o lugar correto, e que é a coisa mais difícil que tem em termo de política pública é porque tem a ver com a colocação de causa e efeito. (COORDENAÇÃO EPR1, grifos da autora da dissertação)

> No primeiro governo Aécio o que se fez foi montar uma estrutura enorme de monitoramento, mas não de avaliação. E quando você fala "estado para resultados" esses resultados, eles não podem ser só entregas, tem que ser o resultado no sentido mais amplo, da eficiência, da eficácia, da efetividade. Então, eu acho que sim, eu acho que o governo sinalizou que ele queria dar um passo adiante, mas eu acho que também se deparou ele, nós nos deparamos, com dificuldades. (AVALIADOR MSF4, grifo da autora da dissertação)

Assim, na ideia do entrevistado EpR1, o governo mineiro já possuía o "M" do monitoramento, por meio de ações de como a implantação do SCGERAES e do acompanhamento de indicadores, no entanto, o "A" de avaliação, ainda constituía uma falha no sistema mineiro de M&A. Além disso, o entrevistado, já antecipando uma preocupação, faz alusão à dificuldade em auferir as relações de causa e efeito das políticas. Nas avaliações que objetivam mensurar o impacto das políticas a maior dificuldade metodológica não consiste em distinguir produtos de resultados, como afirma Arretche (1998), mas em demonstrar precisamente que os resultados encontrados (sejam eles positivos ou negativos) estão causalmente relacionados aos produtos oferecidos por uma dada política sobre análise. Sendo que essas dificuldades, completa Arretche (1998), vão desde as aparentemente mais simples, como a obtenção de informações sobre os programas, até as metodologicamente mais complexas, que dizem respeito à possibilidade de isolar a interferência das variáveis intervenientes em qualquer sistema aberto próprio da análise social.

Outro entrevistado, também da coordenação do EpR, completa o comentário do entrevistado EpR2 sobre as razões de se iniciar a agenda de avaliações no governo de Minas. Além de corroborar a ideia de que havia informações faltantes na, já consolidada, cultura de monitoramento do governo, o entrevistado introduz dois novos motivos que contribuíram para a construção da agenda de avaliações do governo: a necessidade de deter mais informação para a transição do novo governo para fazer uma escolha mais motivada dos projetos efetivos, o que aconteceria em 2011, e a influência do Banco Mundial que mostrava as experiências internacionais de outros governos no campo de avaliação, como exemplifica o entrevistado.

> São três razões. Primeiro, no primeiro governo o fato de ter feito monitoramento intensivo nos dava um conjunto de informações que quase naturalmente pedia um passo à frente, que era avaliações dos programas. Segundo, quando a gente foi fazer o planejamento estratégico do segundo governo, na hora de avaliar os programas para fazer a nova carteira, a gente viu que faltava também um pouco mais de informações além do monitoramento. Depois, o grupo foi melhorando, e, vendo as experiências internacionais, influenciado pelos debates com o Banco. Então, fomos influenciados também pelos organismos internacionais e pela teoria mesmo sobre a gestão pública de qualidade. Então, <u>são essas três coisas: a melhoria, o monitoramento, a necessidade concreta de ter mais dados para fazer uma escolha de um conjunto estratégico no segundo governo e a influência do conhecimento internacional, vamos chamar assim, dos consultores do Banco, da biografia, do grupo que estava ali conosco e que foi melhorando</u>. Mas é aquela ideia, o programa Estado Para Resultado que teve, nós éramos uma unidade de pesquisa em movimento na área de monitoramento e avaliação, nós estávamos ali aprendendo com os consultores do Banco Mundial, trabalhando com gente de vanguarda. (COORDENAÇÃO EPR2, grifo da autora da dissertação)

Assim, de acordo com a percepção do entrevistado, a justificativa do período em que o governo se encontrava também foi estratégica para a introdução das avaliações. A produção de avaliações em um período de transição política, de troca entre governos, disponibilizaria instrumentos de decisão mais técnicos, que capacitariam os tomadores de decisões a exercer uma deliberação mais motivada e, assim, mais adequada aos princípios democráticos. Soma-se ao momento do governo, a presença do Banco Mundial e, a partir daí, do reconhecimento de algumas experiências internacionais que caminhavam na rota das avaliações. E, de fato, vários outros entrevistados citam a presença e a importância do organismo internacional para o *kic-off,* ou seja, para o ponta pé inicial do início das avaliações governamentais em Minas Gerais.

Ferreira, Lopes e Campos (2010) compartilham da idéia de que constituir um conjunto de avaliações, em especial, avaliações de impacto, remete ao fato da necessidade de ter, ao final de um ciclo governamental, um conjunto mais confiável de informações acerca da efetividade das principais políticas públicas no Estado. Ao tentar responde a pergunta "Quais eram as expectativas em relação à avaliação no período em que se decidiu avaliar?", os entrevistados que participaram da formulação do projeto de avaliações relataram perspectivas positivas e variadas com a introdução do pacote de avaliações no governo naquele período. Diferentes tipos de dimensões do uso das avaliações – como já discutido neste trabalho por meio das falas de Leviton e Hughes (1981) e demais autores, podem ser identificados nesse primeiro momento por meio dos relatos dos entrevistados.

A primeira dimensão de uso que pode ser identificada na fala dos entrevistados é a necessidade de promover o uso instrumental que, de acordo com Leviton e Hughes (1981), diz respeito aos casos nos quais as conclusões das avaliações são utilizadas fundamentalmente para a tomada de decisões ou para a solução de problemas dos programas. Ou, como afirmado por Patton (2008), o uso instrumental quando os tomadores de decisões utilizam as informações geradas pelas avaliações para modificar o objeto da avaliação de alguma forma. De acordo com os entrevistados MSF4 e EpR1, a grande expectativa seria poder desenhar e formatar políticas de forma mais técnica, de forma mais segura e mais acertada utilizando as informações contidas nas avaliações realizadas.

> E para mim foi ficando claro que era uma avaliação interna, era o governo avaliando suas políticas para promover melhorias, redesenhar, ver se era caso de persistir, mas não era algo de controle social, que seria exposto para a sociedade. Mas acho que a visão era muito mais de algo interno, um processo de monitoramento e avaliação para uso interno para melhorar os projetos, o desenho, mas não de controle social que as pessoas pudessem dizer "esse dinheiro está sendo bem gasto, está gerando efetividade...", que é uma outra concepção de avaliação de políticas publicas. (AVALIADOR MSF4, grifo da autora da dissertação)

> Então a nossa grande expectativa era ter mais informação. Por quê? Porque eu vi como é mal feita a parte de desenho da política pública. Ou seja, senta seis pessoas à mesa e fala assim: "Vamos desenhar a política". E como você não tem base em monitoramento, não tem informação do que já foi feito e um modo sistemático de se conseguir as informações, você parte de premissas pouco críveis. Então ter essa sistemática de avaliações é importante para você criar maturidade de informação para esse modelo e novos ciclos de desenho, e aí o ideal era de que você criasse uma base de que todas as políticas fossem avaliadas, que tivessem esse registro e que você pudesse a cada ciclo de gestão poder ir aprimorando e tendo a ampliação dessa base servindo não só para você como para outros governos, municipais, outros estados e tudo o mais. E no médio prazo é você ter elementos para você poder desenhar ou criar, adaptar desenhos para novas políticas, que na verdade esse é o grande intuito da avaliação, é o que ela gera de *feedback*, um *feedback* no sentido de adaptar a política para acertar o caminho, acertar mais, acertar menos, ou dedicar mais dinheiro ou dedicar menos dinheiro. (COORDENAÇÃO EPR1)

> Em muitos dos discursos parecia que isso era feito de forma experimental; eles iam fazer de algumas políticas escolhidas para ver como se desenvolveria a avaliação e com grande expectativa de que lá no início gerasse informação para ajudar os gerentes ou os secretários a tomar decisões em relação aos projetos e também que justificasse a manutenção deles ao aumento da quantidade de recursos aplicados nessas avaliações. Então, é isso, era uma tentativa que me foi apresentada assim, dentro do contexto geral do Estado para Resultados de trabalhar em cima de evidências buscando, então, nas avaliações essas evidências do que estava dando certo, do que não estava, do que estava dando resultado e do que não estava. (TÉCNICO EPR3, grifo da autora da dissertação)

Além disso, o entrevistado EpR1 ainda menciona a esperança de que os resultados das avaliações servissem para outros governos e municípios, ele alude um uso que vai além do instrumental, o *enlightenment* ou uso para o esclarecimento. O uso do tipo *enlightenment*, de acordo com Weiss (1998), ocorre quando a avaliação é utilizada para influenciar outras instituições ou eventos que vão além do escopo de determinado programa avaliado. Tal dimensão de uso se justifica uma vez que as conclusões de uma avaliação podem ultrapassar seus limites originais e pautar, não somente os coordenadores de programas avaliados, mas comunidades acadêmicas e coalizões políticas interessadas nos resultados de dada política pública.

Outra dimensão de uso qualificada pelos entrevistados remete à dimensão simbólica ou política, que ocorre quando as avaliações têm a função de legitimar o processo e ganhar

adeptos, servindo como instrumento de persuasão (WEISS, 1998). Tal uso pode ser identificado na fala do entrevistado EpR1 "gerar uma mídia positiva".

> A avaliação tem sempre uma questão que é o curto prazo, ou seja, a necessidade de você gerar uma mídia positiva, um elemento positivo para os seus projetos, com credibilidade que você pudesse fazer isso, sabendo o risco de que poderia ser uma mídia negativa. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

Outra dimensão de uso que pode ser identificada relaciona-se ao uso persuasivo (PATTON, 2008), à utilização seletiva dos resultados da avaliação para dar suporte às decisões específicas sobre financiamento e debates políticos. Para Arretche (1998), a criação e o fortalecimento de estudos confiáveis de avaliação permitem um uso mais consciente dos recursos na medida em que as informações dos estudos são utilizadas para evitar o desperdício de recursos públicos. De acordo com o entrevistado EpR2, o objetivo mais ambicioso das avaliações seria orientar a alocação de recursos no governo. No entanto, como o próprio entrevistado completa, "essa é uma expectativa muito alta e muito difícil".

> Talvez o resultado mais efetivo seria orientar a alocação de recursos para aqueles programas que tivessem produzido maior resultado para os seus beneficiários. Essa é uma expectativa muito alta e muito difícil de se fazer dado as vinculações funcionais que nós temos. Nunca tive a ilusão de, no estágio que nós estamos, dado a vinculação orçamentária que existe na legislação brasileira, tem um orçamento orientado por resultados. (COORDENAÇÃO EPR2, grifo da autora da dissertação)

E por fim, apresenta-se o uso processual, cunhado por Patton (2008), como o uso que ocorre quando pessoas envolvidas nas avaliações, com exceção dos avaliadores, demonstram alterações comportamentais e cognitivas, como resultado da participação no processo de avaliação. As mudanças pretendidas pelo uso processual são reflexo, portanto, da aprendizagem experimental adquirida ao longo da realização de uma avaliação e da convivência com avaliadores mais experientes, apreendendo seus valores e forma de raciocínio. Para o entrevistado da coordenação da FJP, uma ideia ambiciosa seria tornar a FJP um centro de referências quanto à avaliação de políticas, de forma que o conhecimento adquirido nas avaliações fosse absorvido não só pelos avaliadores do processo, como pelos alunos da Escola de Governo[5] que, por sua vez, transmitiriam à ideia ao governo estadual. Para o entrevistado do EpR, o processo também geraria conhecimentos para o próprio grupo da instituição que coordenava as avaliações, além de um aprendizado para o Centro de Desenvolvimento e Planejamento Regional da Universidade Federal de Minas Gerais

[5] Para melhor entendimento da lógica entre a Escola de Governo da FJP, ver Oliveira (2008).

(CEDEPLAR/UFMG), outro ente que participou das avaliações educacionais como poderá ser verificado mais a frente:

> A ideia era ambiciosa. Na verdade, na cabeça da gente era de que você teria a equipe técnica da Fundação João Pinheiro cada vez mais experiente na condução nesse tipo de projeto e os alunos da Escola de Governo seriam estagiários no projeto, e depois isso funcionaria como mecanismo de difusão da cultura de avaliação de monitoramento no governo estadual. Depois que os alunos saíssem de lá, fossem para os órgãos, tendo tido essa experiência eles seriam os veículos, os transmissores, os difusores dessa cultura de avaliação pela estrutura toda. Nada disso aconteceu, mas era essa a ideia. (COORDENAÇÃO FJP)

> Para mim era um processo de aprendizado inclusive do grupo do Projeto Estado para Resultados, que era um grupo novo. Então, assim, era um aprendizado do programa Estado para Resultados, um aprendizado para a Fundação João Pinheiro, um aprendizado para o Cedeplar. Que era um pouco de criar uma rede em Minas Gerais que tivesse o conhecimento de avaliações, porque nós já tínhamos avançado bastante no monitoramento no Estado com o *status report*, com toda experiência do GERAES, então seria interessante isso. E a segunda expectativa era produzir um resultado para dar um retorno para área setorial e eventualmente orientar uma alocação de recursos. (COORDENAÇÃO EPR2, grifo da autora da dissertação)

Dessa forma, observa-se que, na visão dos entrevistados, as avaliações que estavam sendo contratadas naquele momento motivavam expectativas ousadas de diferentes tipos de utilização futura. É possível verificar a presença, não somente do uso instrumental, que culmina na tomada de decisões a partir dos resultados da avaliação, estava sendo buscado, mas também outros quatro usos secundários, como a utilização do tipo *enlightenment,* que vai além da alçada do projeto avaliado, do uso simbólico ou político, da dimensão persuasiva e, por fim, um anseio imenso pelo uso processual. Weiss (1998), no entanto, destaca que, para alcançar um uso instrumental, é necessário que algumas condições mínimas sejam atendidas como, por exemplo, que as conclusões das avaliações não sejam controversas, nem provoquem conflitos de interesse na estrutura da organização; que as mudanças a serem implementadas se encontram dentro do escopo de atuação do programa e não provoquem grandes rupturas no funcionamento do programa; e que o ambiente do programa seja relativamente estável, não apresentando dúvidas quanto ao seu orçamento, público-alvo e implementação entre outros. Observam-se, nesse sentido, as várias possibilidades de uso que podem ser geradas por avaliações e, ao mesmo tempo, a dificuldade de alcançar o uso instrumental, dado que uma série de circunstâncias devem ocorrer para que a mudança intrínseca à conclusão de uma avaliação aconteça. De acordo com o entrevistado do EpR, os governos não usam todas as possibilidade que uma avaliação gera, sendo a maior dificuldade conseguir mudar o curso de gestão em função dos resultados obtidos. Para o técnico do EpR,

o uso deve ser incentivado e insistido para que governos não corram o risco de investir onde não há impacto.

> A gente não usa, no meu modo de entender, toda a possibilidade que uma avaliação gera. A dificuldade é mudar o curso da gestão em função desse resultado. (COORDENAÇÃO EPR2, grifo da autora da dissertação)

> Acho que a gente pode estimular mais a utilização dessas informações, insistir mais, é o que a gente não tem feito também. Acho que a gente tem que insistir nisso e continuar fazendo, porque senão a gente vai continuar correndo o risco de investir errado, investir onde não deve, onde não dá resultado, enfim... (TÉCNICO EPR1, grifo da autora da dissertação)

Desenvolvidos os passos iniciais do movimento que culminaram nas avaliações, constrói-se, a seguir, a partir do relato dos entrevistados, a implementação e o funcionamento para a execução das avaliações. Destacam-se temas como o treinamentos dos avaliadores, a seleção dos projetos que sofreriam a avaliação e o envolvimento dos gerentes políticos e administrativos das áreas setoriais.

6.2 Processo geral de implementação e execução das avaliações

O primeiro esforço para a implementação das avaliações, já após o fechamento do contrato com o Banco Mundial, em 2008, foi o início de um programa de treinamento para técnicos da FJP e do EpR. Ferreira, Lopes e Campos (2010) contam que o marco para o início da agenda de avaliações no governo mineiro foi a promoção, em 2007, de um curso pelo Banco Mundial sobre avaliação de impacto que contou com a participação de especialistas do EpR e, a partir disso, e, com encontros subsequentes, com especialistas do Banco Mundial em Minas Gerais, iniciou-se a identificação de alguns projetos que poderiam ter uma avaliação de impacto realizada. A primeira preocupação da equipe da FJP foi o treinamento dos funcionários e, posteriormente, o desenvolvimento de um projeto de avaliação mínimo para os projetos selecionados. De acordo com o entrevistado da FJP, começou uma grande ação de treinamento da equipe técnica da FJP que iria trabalhar nas avaliações, contando com o auxílio de especialistas nacionais e especialistas do Banco Mundial no tema de M&A.

> Primeiro, começou um grande programa de treinamento da equipe técnica da Fundação João Pinheiro para a avaliação de políticas. Trouxemos vários professores de outras instituições para ministrar cursos para os técnicos da Fundação. Com a ajuda do Banco Mundial a gente trouxe também algumas pessoas aqui. Não vou me lembrar agora de tudo o que a gente fez, mas uma parte do esforço consistiu nisso,

> treinar a equipe técnica da fundação nas novas técnicas de avaliação de políticas públicas. (COORDENAÇÃO FJP, grifo da autora da dissertação)

> No início a gente teve alguns *workshops* com o pessoal do Banco Mundial, que veio fazer uma espécie de treinamento com a equipe, a gente participou também de cursos da Fundação Itaú Social, que são cursos de Avaliação Econômica de Impacto. E a equipe também procurou sempre ler e buscar literaturas a respeito de avaliação de políticas públicas. Na época também a gente teve aqui um congresso da rede de avaliação e monitoramento que também foi um ponto positivo. Para a Fundação e para os membros da equipe foi algo bem positivo participar desse tipo de avaliação. A Fundação, ela já te uma trajetória de avaliação, de parcerias, mas para o tipo de avaliação que foi solicitado eu acho que foi interessante porque a equipe aprendeu muito também. Eu acho que foi uma oportunidade boa. (AVALIADOR VV1)

> Pra isso houve um curso aqui com o Bamberger sobre avaliação de impacto, a gente foi incentivado a fazer um curso na Fundação Itaú Social. Então, de alguma forma, a equipe foi preparada: seminários. (AVALIADOR VV2)

Além da preocupação em criar *expertise* quanto à avaliação, a própria FJP persistiu na contratação de força de trabalho qualificada para a concussão das avaliações, que se somaria ao corpo de funcionários públicos de carreira da FJP. De acordo com o entrevistado da FJP, a intenção era contratar profissionais qualificados, todos com título de doutorado, selecionados por um comitê técnico, de forma semelhante ao que acontecia com o processo de contratação dos empreendedores públicos no estado. Além disso, o contrato desses profissionais seria temporário, com fim definido em janeiro de 2011.

> A gente montou equipe com os técnicos que já estavam lá e com pessoas novas que se levou por meio de vários mecanismos: concurso público que se fez e o cargo de Gestor de Ensino e Pesquisa que foi criado para poder contratar pessoas com doutorado para a Fundação João Pinheiro. Um cargo de confiança foi criado. A inspiração, a ideia foi o setor público criou 30 cargos de gestão de ensino e pesquisa na Fundação João Pinheiro, o requisito para ocupar o cargo era ter doutorado, a seleção era feita por um comitê também. Havia uma pré-qualificação por um comitê à semelhança do que acontece no Estado para Resultados, o presidente da Fundação João Pinheiro nomeava, mas com essa restrição, só podia nomear quem tivesse sido pré-qualificado por esse comitê. Era um cargo que ia acabar em janeiro de 2011, aquela coisa toda. Mas fez o concurso público para carreira na Fundação e seguiu esse cargo também, e se recrutou pessoas bem qualificadas. (COORDENAÇÃO FJP)

Parte da estratégia de capacitação e desenvolvimento de competências de avaliação nos funcionários da FJP, tanto dos novos funcionários temporários como dos funcionários de carreira, contou com a criação da figura de um *consultor* contratado especificamente para ajudar no desenvolvimento de metodologias próprias para cada projeto avaliado. De acordo com o entrevistado da FJP, a ideia de contratar consultores faz parte do planejamento *maior*

de criar o aprendizado quanto ao tema avaliação e criar certa internacionalização da tecnologia.

> E a estratégia que a gente montou foi de contratar para cada um desses projetos um consultor externo com muita experiência na área. A ideia era de que esses consultores colaborariam com a equipe da Fundação João Pinheiro também com aquela ideia de fazer a internalizarão da tecnologia. Eram especialistas importantes na área no Brasil e o papel deles eram um pouco esse, colaborar na formulação dos projetos, supervisionar a elaboração.... Então a ideia era de que esse era o mecanismo de internalizar a tecnologia, de importar a tecnologia. Esses consultores colaborariam na formulação do projeto, supervisionariam o andamento do trabalho. Mas o trabalho seria feito pelas equipes da Fundação João Pinheiro. (COORDENAÇÃO FJP)

> Os cursos que a gente fez foram importantes, mas eu acho que não o suficiente para a gente já ir partindo para uma avaliação de impacto sem nenhum apoio. Então, as avaliações contaram com especialistas de renome nacional. Eu acho que a gente não se atreveria a formular todo o desenho da avaliação sem um apoio. (AVALIADOR VV2)

Esse discurso é corroborado por outro entrevistado EpR1 quando ele afirma que, mesmo com a contratação dos doutores, a FJP ainda não tinha habilidade para entender todas as dificuldades e limitações próprias da natureza das avaliações de impacto. A *figura do consultor*, como afirmado pelo entrevistado EpR1, nasceu para desenvolver a capacidade dentro da própria Fundação de modo que a experiência pudesse ser replicada para projetos futuros. Para o entrevistado, os consultores, por ser um caráter temporário, deveriam nortear e orientar as equipes de avaliadores para tornar o aprendizado o melhor possível.

> A figura do consultor nasceu para desenvolver capacidade. Ou seja, sabia-se que não se tinha capacidade instalada na Fundação e que mesmo trazendo os doutores e tal, eles dominavam essas técnicas, digamos. E o presidente da Fundação, um olheiro muito iluminado, falou assim "o mais importante é desenvolver capacidade". Não é contratar o consultor, mandar ele fazer a avaliação e ele entregar, mas, sim, contratar o consultor, mandar ele sentar com a equipe e desenvolver a capacidade dentro da Fundação para que ela pudesse refazer, replicar. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

A participação dos consultores também se justificaria, segundo os entrevistados, pela visão isenta quanto aos projetos do governo mineiro. Para a realização das avaliações, seria necessária a participação de especialistas externos ao governo, tanto para a aplicação de métodos inerentes à avaliação cuja expertise ainda não era desenvolvida no governo, quanto para a condução de avaliações mais imparciais. A ideia seria, portanto, acionar avaliadores alheios à organização e que, assim, como afirmado por Aguilar e Ander-Egg (1995), tinham maior objetividade na consecução das avaliações. Para o entrevistado EpR2, uma das

preocupações iniciais seria que as avaliações não fossem influenciadas pelos grupos de interesse envolvidos.

> Foi uma escolha intencional, quer dizer, trazer alguém com um olhar mais isento. Eu acho que no caso da Avaliação de Impacto isso tem que ser uma preocupação, porque senão já desde o viés do interesse, do grupo de interesse já vem... A avaliação de impacto, ela poderia ser tendenciosa desde o início, desde a montagem. Então, a ideia de trazer terceiros é essa. A própria Fundação João Pinheiro tem uma terceira unidade que não estivesse ligada nem a unidade central nem a unidade setorial. Tentar isolar a visão dos grupos, né. (COORDENAÇÃO EPR2)

E assim, a partir de treinamentos, intercalados com encontros realizados entre especialistas do Banco Mundial e técnicos do EpR, foram selecionadas os primeiros projetos que receberiam avaliações de impacto ainda em 2008 (FERREIRA; LOPES; CAMPOS, 2010). Na visão de diferentes entrevistados, contudo, a seleção dos projetos estruturadores que seriam avaliados contou com uma escolha feita de modo arbitrário, mais ou voltada para certas áreas consideradas mais relevantes. Quando interrogados pelos critérios de seleção dos projetos que passariam pelas avaliações de impacto, três entrevistados mencionaram que a justificativa das avaliações se encontrava na importância relativa da área em que o projeto se encontrava, ou seja, a opção pelos projetos, nesse sentido, deu-se pela opinião do grupo que estava selecionando de quais os setores do governo que mereciam uma avaliação.

> Não tem uma fórmula especifica assim, que é matemática, mas basicamente você seguiu a ideia: 1- é estratégico, e do estratégico, partiu do pressuposto é projeto estruturador, e está em áreas com um destaque maior. Porque a gente sabia também que nem todos os estruturadores tinham o mesmo nível de importância, então, no ponto de vista estratégico, você tinha uma sensibilidade de que algumas áreas como educação, saúde e segurança são mais importantes do que outras áreas. Ou seja, importância no ponto de vista que da setoriedade de governo e do governante, ou seja, que eram mais o queridinho dos olhos. (COORDENAÇÃO EPR1)

> As áreas eram mais ou menos óbvias. Porque área de educação, saúde e segurança eram mais ou menos óbvias. Então, é mais ou menos óbvio que dada a importância dessas áreas, que os primeiros projetos de avaliação se referissem a programas nessas áreas. A gente quis também pôr alguma coisa na área de política social, acho que provavelmente isso foi importante. (COORDENAÇÃO FJP)

> Então, assim, nós tínhamos conhecimento tácito e objetivos, desde o início, eu como subsecretário e ele como superintendente lá, que depois virou o GERAES, a gente tinha um conjunto de informações sobre os programas. Então, a gente via aqueles programas que tiveram maior importância no primeiro governo, que eram melhores estruturados e que talvez tivessem uma capacidade de produzir um resultado da avaliação melhor. Então, foi esse grupo que escolheu, com base nas informações do monitoramento, com base na escolha do segundo governo. Então, assim, uma escolha estratégica desse grupo. (COORDENAÇÃO EPR2)

Assim, a área educacional foi a primeira a se destacar. Foram selecionados quatro projetos estruturadores: Poupança Jovem, Escola em Tempo Integral, Aceleração da aprendizagem e Programa de Educação Profissional. Para isso contou-se com a participação de núcleos de pesquisa como o CEDEPLAR (FERREIRA; LOPES; CAMPOS, 2010). De acordo com o entrevistado EpR1, a escolha do CEDEPLAR deu-se pela experiência da instituição na área educacional e pela importância do setor para o governo.

> O CEDEPLAR veio por uma questão de que a educação tem muita informação, tem uma política muito ousada de base de dados, era um apelo muito grande do Governo do ponto de vista estratégico, e aí tentou fazer uma coisa mais ousada com a educação no sentido do pacote da avaliação. (COORDENAÇÃO EPR1)

Em seguida, outros quatro projetos foram selecionados para passar por avaliações: dois programas de destaque na área de Saúde, Viva Vida e Pró-Hosp; e um projeto de segurança pública, Gestão Integrada de Ações e Informações do Sistema de Defesa Social. A escolha do projeto Minas Sem Fome, de acordo com o entrevistado da FJP, se justificou pela necessidade de possuir um projeto particular da área de redução da pobreza e pela importância relativa do projeto à época.

O entrevistado EpR3 não descarta a questão política no momento de se selecionar os projetos para a avaliação. Além de citar que a escolha não foi essencialmente técnica, o entrevistado, aponta que os projetos selecionados contavam com orçamentos vultuosos e havia certa expectativa que a relação entre causa e efeito desses projetos era positiva, ou seja, esperava-se que os projetos gerassem resultados concretos para a sociedade. Nesse sentido, o palpite do entrevistado EpR3 é que a escolha dos projetos tenha sido feita com o intuito de justificar e afirmar que a estratégia do governo estava certa, que os projetos avaliados por avaliações de impacto suscitariam impactos positivos.

> Acho que foi uma escolha discricionária de pensar dentro da carteira de projetos disponíveis, quais seriam mais interessantes de avaliar tanto pelo orçamento que estava vinculado a cada um deles ou pela expectativa do que já tinha, de que aqueles projetos davam resultados, e aí seria não só a fonte de informação importante para manutenção desses programas para o benefício do cidadão, mas também para justificar as escolhas que o Governo estava fazendo na época. Então imagino que foi isso pensaram. "Ok, esses eu acho que precisam ser avaliados porque envolve um orçamento muito grande ou trabalham numa área na qual o impacto pode ser bastante relevante, e, além disso, se eles estiverem dando resultado e a gente tiver como comprovar isso, ganhamos respaldo político". (TÉCNICO EPR3)

No entanto, o entrevistado EpR1 apresenta um ponto de vista diferenciado. Ao contrário de representar uma decisão política, o principal ponto de escolha dos projetos remeteria a uma seleção técnica. Para o entrevistado, a realização das avaliações os projetos demandariam requisitos mínimos para auferir o impacto como, por exemplo, uma relação causal bem definida. Assim, alguns projetos que, simplesmente, não apresentavam uma relação causal bem definida não poderiam ser selecionados.

> Mas não só isso, aí juntava o segundo ponto, que era o principal, que era a capacidade de avaliar. Você tem muitos projetos, por exemplo, de desenvolvimento econômico que você não tem como fazer Avaliação de Impacto. Por quê? Porque, primeiro, não tem causa e efeito. Com relação de causa e efeito elas são muito mais difíceis. Segundo, não se permite apurar essas avaliações de causa e efeito. Alguns projetos você queria avaliar, mas não tinha capacidade de avaliar, então você abortava, deixava, não escolhia ele para escolher um outro. (COORDENAÇÃO EPR1)

Alem disso, o entrevistado EpR1 ressalta o papel dos consultores para um segundo *filtro* dos projetos avaliados. Haveria uma primeira seleção por parte das pessoas que selecionaram os projetos e, posteriormente, os consultores refinavam essa escolha por meio da checagem de requisitos mínimos para a realização de uma avaliação de qualidade. Nesta hipótese, os projetos que se enquadravam nos requisitos de uma avaliação de impacto, poderiam ser selecionados.

> Só com base no que a gente sabia do acompanhamento que a gente tinha dos projetos a gente já falava "Esse aqui dá, esse aqui não dá". Obviamente o consultor ia perguntando "Tem isso? Não tem isso?" e depois a gente fazia um comando com o gestor e avançava para avaliar, e aí descobria o que de fato tinha, o que não tinha, descobria algum outro dado a mais ou a menos. (COORDENAÇÃO EPR1)

De acordo Aguilar e Ander-Egg (1995, p. 69) avaliar um programa ou projeto não é algo que possa ser feito em qualquer tempo e lugar, isto é, "[...] não é algo que se possa fazer simplesmente porque se decide fazê-lo". É necessário, na visão dos autores, que, além de requisitos lógicos e científicos, que os responsáveis políticos e administrativos pelos programas estejam plenamente convencidos de que a avaliação é um caminho necessário. A decisão de avaliar foi realizada por um grupo informal composto pela alta direção da SEPLAG, especificamente pelo subsecretário responsável pelo SCGERAES, pela coordenação executiva do EpR e pelo presidente da FJP. O grupo compunha um comitê de coordenação da agenda de avaliações que, de acordo com o relato de entrevistados, se reunia regularmente e se responsabilizava pelo gerenciamento dos cronogramas de avaliação. De

acordo com o entrevistado EpR1, o comitê era uma maneira *institucional* e menos personificada de selecionar os projetos das avaliações.

> Porque até então tentou-se criar uma maneira institucional de fazer isso [seleção dos projetos a serem avaliados], tirar a questão de pessoa, então criou-se um comitê de avaliação, que era composto pela Fundação, representantes do Estado Para Resultados e de representantes acho que da SEPLAG... (COORDENAÇÃO EPR1)

Na visão de outro entrevistado, no entanto, apesar de reconhecer que o comitê seria o ente responsável pela coordenação das avaliações, tem dúvidas sobre a efetividade das ações do órgão colegiado. E o próprio entrevistado EpR1 se contradiz sobre as ações de gerenciamento do comitê. Segundo o entrevistado, o nível de gerenciamento de projetos não seria o mais detalhista.

> Tinha um comitê que eu não lembro como é que chama que supostamente seria a coordenação dessa coisa toda. O comitê se reunia, mas era uma coisa um pouco mais formal. Eu não acho que o comitê mesmo foi muito efetivo na coordenação disso. (COORDENAÇÃO FJP)

> A gente acompanhava, reportava nesse comitê. Obviamente, não vou falar que era um nível de gerenciamento de projeto primata, porque eu acho que não precisa no sentido de nível de detalhe e tal. (COORDENAÇÃO EPR1)

Apesar do comentário de Aguilar e Ander-Egg (1995), de que os gestores dos programas a serem avaliados devem estar cientes sobre os propósitos, aplicação e possíveis consequências da avaliação, tais atores só começaram a participar das avaliações mineiras quando foram informados sobre a ideia tomada pelo comitê de avaliar seus projetos. Patton (2008) reitera que além de se comprometerem com a decisão de avaliar, os gestores políticos e administrativos têm que aceitar e incluir a avaliação como parte integrante de seus projetos. Observa-se, no entanto, que a seleção da agenda de avaliações do governo não contou com a participação dos gestores políticos e administrativos dos projetos no primeiro momento.

Quanto à participação das equipes dos projetos avaliados no planejamento e implementação, o entrevistado do Cedeplar e o técnico do EpR abordam a dificuldades de coordenação do EpR. De acordo com Bamberger, Rugh e Mabry (2011), o que distingue um estudo avaliatório útil de um inútil é a extensão de envolvimento dos usuários ao longo do processo, para isso, os clientes devem se sentir *donos* das avaliações e acreditar nas informações geradas pelas mesmas possa agregar valor ao seu trabalho, o que nem sempre ocorreu nas avaliações contratadas pelo EpR. A relação entre os interessados em avaliar, EpR e FJP, e aqueles que

teriam seus projetos avaliados, integrantes das secretarias, nem sempre foi tranquila como apontado por algumas entrevistas. Nesse sentido, a legitimidade do EpR para negociar, organizar e executar as avaliações foi questionada, tanto pela dificuldade de penetração nas unidades setoriais, quanto pelo desconhecimento das necessidades informacionais de cada projeto, como exemplificado nas entrevistas.

> Eu acho que um pouco essa coisa de como é que você administra dentro do Governo. O Estado para Resultados tentou integrar esse tipo de coisa, mas ele tinha as dificuldades porque alguns secretários nem aceitavam conversar. Então, assim, eu acho que essa tentativa do Estado para Resultados é importante, mas tinham problemas desse tipo: como é que você conversa com secretarias? Como é que você pega o comprometimento deles de que a política vai ser avaliada passando por ali? Como é que a Secretaria de Estado para Resultados trabalha junto com as outras secretarias? Fica parecendo que está intervindo. Então você tem várias dificuldades. Não sei se isso é fraqueza ou ponto forte, acho que foram dificuldades. A ideia de ter uma secretaria para isso é muito boa, mas aí logo aparecerem todas essas dificuldades. (COORDENAÇÃO CEDEPLAR, grifo da autora da dissertação)

> Acho que o próprio Estado Para Resultados estava muito, muito focado no impacto, uma obsessão por avaliação de impacto. E acho que pouco conhecimento de avaliação até para pensar qual que seria a avaliação. Ele tinha que está mais aberto para pensar o que avaliar e como avaliar. Acho que principalmente isso. E pensar sempre: o que avaliar, você não pode fazer isso sem... Essa definição tem que virou da secretaria ou mesmo que seja uma definição do governador, sei lá. Mas tem que está muito claro "Por que eu estou avaliando esse programa?". Eu acho que o estado Para Resultados não tinha legitimidade para definir o que deveria ser avaliado. (TÉCNICO EPR1, grifo da autora da dissertação)

A falta de interação entre os avaliadores e a própria equipe dos projetos avaliados gerou a percepção de que o cliente das avaliações seria o EpR, ao contrário dos gestores dos projetos avaliados, o que, na visão do entrevistado do EpR, seria uma anomalia da avaliação, como afirmado no trecho abaixo. Para o entrevistado, como o cliente original (o gestor) nem sempre entendia o processo da avaliação, não tinha tempo para dedicar-se ao assunto e nem proatividade, o EpR assumiu a responsabilidade de entregar o conjunto de avaliações para o Banco Mundial.

> O cliente da avaliação era o Estado para Resultados. Então você tinha um intermediador nesse processo que não é o ideal. O ideal é que o cliente fosse o gestor. Como o cliente muitas vezes não entendia o processo, como o cliente muitas vezes não tinha tempo para se dedicar ao processo, como o cliente muitas vezes o cliente, o gestor, não tinha proatividade para tocar o processo, essa figura intermediaria assumiu mais do que talvez deveria para garantir uma entrega, afinal de contas você tinha que entregar porque isso valia o Banco Mundial. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

O diálogo com os gerentes das áreas que seriam avaliadas, de acordo com o entrevistado EpR1, era tenso. A partir do momento em que um grupo reduzido de pessoas seleciona os projetos que passariam por avaliações, seria necessário que, em algum lugar da história as partes avaliadas fossem minimamente informadas. E isso aconteceu. No entanto, como afirma o entrevistado EpR1, já existia certa resistência das áreas setoriais por instrumentos de controle dado a experiência de monitoramento dos projetos realizada pelo SCGERAES, e a presença de outras instituições realizando as avaliações acirrava essa percepção. No comentário a seguir, observa-se que o gerente político do projeto, geralmente o secretário da pasta, era o que apresentava mais oposição. Para o entrevistado EpR2, uma unidade de avaliação central de governo sempre será vista como uma instituição de controle como demonstram os relatos abaixo.

> E o momento de você sentar com o gerente era um momento tenso, como todo momento de avaliação, que, primeiro, deixar claro o que você está avaliando, e aí o problema nem era o gerente, o geralmente o problema era o secretário, porque você tinha interferência numa pasta de outra pasta. Ou seja, esse já era o rescaldo do problema do GERAES, que era ter muita força o planejamento e menos no finalístico e que você estava sofisticando ainda mais e levando um custo maior no sentido de prestar mais informação, fazer mais coisa. Porque no fundo a cultura de avaliação é isso. Então o que a gente tentou fazer foi mostrar a importância da avaliação. E trazer gente do Banco Mundial e tentando fazer as pessoas entenderem isso e se sentirem, digamos, tão mais donas do processo do que a gente. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

> Isso talvez tenha sido uma deficiência nossa. A gente quando está no órgão central, a gente acaba olhando do planejamento, vendo o conjunto, aí depois você vai conversar um pouco com a área finalística. Porque normalmente se a unidade de avaliação está numa unidade central ou mesmo numa unidade independente, ela é vista como uma unidade controle. Ela sempre vai ser vista como unidade controle. Ao invés de ser vista como uma unidade que está gerando informações de qualidade para tomadas de decisões dos executivos, ela é vista como uma unidade de controle. (COORDENAÇÃO EPR2, grifo da autora da dissertação)

Guba e Lincoln (1989) levam em consideração o contexto em que as avaliações estão inseridas. De acordo com os autores, a desconsideração das circunstâncias e dos propósitos que envolveram a implantação e execução dos programas são problemas comuns que podem comprometer as avaliações. Além disso, Shulha e Cousins (1997) destacam a importância da estrutura e dos processos organizacionais em que as avaliações são colocadas que podem influenciar diretamente, em menor ou maior grau, a transmissão das informações geradas pelas avaliações e a inserção de mudanças no nível organizacional. No caso de Minas Gerais, observa-se que a desconsideração da percepção dos gerentes políticos e administrativos dos projetos e a desconsideração do contexto específico que se apresentava em que cada um

desses projetos poderia causar danos no desenvolvimento das avaliações. No momento em que já existia certa resistência quanto aos instrumentos de controle utilizados pela SEPLAG, a avaliação poderia ser vista como mais uma barreira às atividades das equipes dos projetos.

Além disso, embora o escritório central de projetos do governo, o SCGERAES, desempenhasse um papel essencial no desenho dos projetos estruturadores, oferecendo um planejamento especial para essa carteira, existiam diferenças significativas entre os projetos dependendo da iniciativa e da vontade e capacidade da equipe de projetos (FERREIRA; LOPES; CAMPOS, 2010). Assim, embora alguns projetos fossem selecionados para a realização de avaliações de impacto, eles apresentavam diferenças, tanto quanto ao seu planejamento inicial, quanto a cultura organizacional de cada uma das secretarias de Estado em que os projetos era desenvolvidos, dada a heterogeneidade dos processos de idealização e estruturação dos projetos. Dessa forma, como observado por Ferreira, Lopes e Campos (2010), nem todos os projetos estavam *preparados* para *receber* avaliações de impacto, uma vez que não dispunham de análises anteriores quanto à robustez do seu desenho.

Uma limitação geral das avaliações observada por Ferreira, Lopes e Campos (2010) diz respeito à qualidade dos registros administrativos e dados disponíveis dos projetos avaliados. Como dito pelos autores, ao longo das avaliações de impacto, identificou-se que alguns projetos avaliados não tinham todas as informações necessárias à realização de uma avaliação de impacto. De acordo com o entrevistado da coordenação do EpR, o principal gargalo das avaliações estava na ausência de dados dos projetos para testar os métodos. O grande desafio era ter uma base de dados confiável e acessível para a realização dos estudos, como se comenta no trecho abaixo.

> O principal gargalo em termo de avaliação não é a metodologia de avaliar para fazer a continha, mas é você ter os dados. Uma vez que você tem os dados, que tivesse esses dados de uma maneira sistemática o tempo todo, chamar alguém, pegar os dados, fazer uma avaliação e entregar o relatório é rápido, que foi o que o consultor fez. O problema é ter essa base organizadinha e tal, que é o grande desafio. Ou seja, o melhor dos materiais está nos registros internos e que o Governo não tem capacidade e sistemática de acompanhar. O GERAES, ele fez muito registro interno próprio, mas o registro interno que era muito gerencial e muito focado no orçamento, não era focado em resultado. Então, não se tinha base de dado adequada e, obviamente, uma vez não registrado você perde essa informação. Recuperar é muito mais difícil. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

As perguntas que ficam após todo esse movimento de incentivo à realização de avaliações de impacto das políticas públicas, de treinamento, contratação de centros de pesquisa,

deslocamento de equipes, desgaste institucional, são: vale o tanto que se investiu? As avaliações foram realmente utilizadas a fim de gerar modificações nos projetos avaliados? Tentar responder a essas perguntas é um dos objetivos deste trabalho. Nos capítulos que se seguem, serão apresentados os casos individuais de avaliação dos quatro projetos selecionados: Viva Vida, Gestão Integrada de Ações e Informações de Defesa Social, Minas Sem Fome e, por fim, Acelerar para Vencer, respectivamente.

7 AVALIAÇÃO DO PROJETO VIVA VIDA

A avaliação do projeto estruturador Viva Vida constitui uma das avaliações do pacote organizado pela FJP e EpR em conjunto com o Banco Mundial e, portanto, contou com a participação de consultor contratado e avaliadores da FJP. Para a pesquisa empírica deste trabalho, contou-se com a participação de sete entrevistados específicos da avaliação do projeto, o que incluí: gerente do projeto Viva Vida à época em que foi negociada a avaliação, dois técnicos da Secretaria da Saúde da equipe do projeto, três avaliadores da FJP que participaram em diferentes etapas do processo de avaliação e o técnico do EpR que acompanhou o projeto de avaliação. Também incluem-se percepções de entrevistados da coordenação do EpR e da coordenação da FJP. Torna-se importante ressaltar, de início, que a avaliação do projeto Viva Vida, em comparação com os casos aqui pesquisados, foi a que contou com o maior número de entrevistas e também o maior número de percepções contraditórias sobre o desfecho da avaliação como poderá ser visto adiante.

O capítulo está organizado em quatro seções que remontam o histórico da avaliação do MSF, a partir da percepção dos entrevistados, desde a concepção da avaliação e expectativa inicial, até a sua conclusão, com alusão ao tipo de utilização ocorrido. Na primeira seção apresenta-se o contexto em que a avaliação foi inserida, seguido pela implementação e envolvimento dos atores na execução da avaliação. A terceira e quarta seções tratam, respectivamente, dos resultados e da utilização da avaliação.

7.1 Contexto geral da avaliação

O Projeto Viva Vida adveio do estudo *Análise da mortalidade infantil e materna no estado de Minas Gerais*, um diagnóstico realizado em 2004, cujo conteúdo evidenciou a fragilidade na rede assistencial de atenção à gestante, ao parto e ao recém-nascido, demonstrando que as principais causas dos óbitos infantis estão concentradas no período neonatal, principalmente até o 6º dia de vida. Além disso, o estudo identificou elevadas taxas de morte materna, de acordo com parâmetros da Organização Mundial de Saúde (OMS). Comprometer-se, portanto, com a redução da mortalidade materna-infantil foi um desafio assumido pelo projeto.

Em linhas gerais, o Projeto Viva Vida procurou investir na estruturação de uma rede, cunhada de Rede Viva Vida, em qualificação e mobilização social, com o objetivo de reduzir a mortalidade infantil e a materna em Minas Gerais. A Rede Viva Vida é composta por três níveis de atenção à saúde: atenção primária que é aquela prestada no nível mais próximo da comunidade, via Unidade Básica de Saúde (UBS) e equipes do Programa Saúde da Família (PSF); atenção secundária, que presta atendimento especializado em Centro Viva Vida distribuídos em unidades regionais; e atenção terciária à saúde, que diz respeito à assistência em hospitais regionais com maior nível de especialização. A qualificação da rede é baseada, principalmente, no Plano Diretor da Regionalização (PDR), uma proposta de organização dos serviços de saúde em 10 oficinas que são realizadas nos municípios e microrregiões. E, por fim, a mobilização social tem como objetivo inserir representantes de organizações governamentais e não governamentais, que são estimulados a agir em defesa da redução da mortalidade infantil e da materna, como exemplo dos comitês estaduais e municipais de Defesa da Vida.

A estruturação da Rede Viva Viva diz respeito à implementação de Centros Viva Vida (CVV), localizados nas microrregiões mineiras, e com o objetivo à assistência integral à saúde sexual e reprodutiva das mulheres, homens e à criança de risco. O CVV atua no planejamento familiar e atenção às gestantes, parturientes, puérperas, recém-nascidos e as crianças até um ano de vida. É importante ressaltar que os CVVs funcionam como fonte de atenção secundária à saúde. Assim, o serviço especializado não deve ser procurado diretamente pelo usuário, mas, sim, ser referenciado pela atenção primária (UBS ou PSF). Para o gerente do projeto, a estratégia dos CVVs constituía, à época, uma proposta pioneira frente ao modelo do Sistema Único de Saúde (SUS).

> A proposta em relação ao centro, diferente de tudo que o estado tinha feito até então, não era só arcar com o investimento, era realmente fazer uma diferença, porque a média complexidade sempre foi um vazio no SUS. (GERENTE DO PROJETO VV)

Para os entrevistados, vários motivos colocavam o projeto Viva Vida na *vitrine* da Secretaria da Saúde e contribuíram para sua escolha para passar pela avaliação de impacto. Os motivos vão, desde a sua proposta única e diferenciada, que envolve várias esferas do sistema de saúde, até a sua vinculação como o planejamento estratégico do governo e com a meta de redução da mortalidade infantil no estado. De acordo com o entrevistado VV2, o desenho

coerente do projeto e lógica de atuação do projeto contribuíram para que ele fosse selecionado.

> Imagina, você tem um plano de governo de tornar melhor estado para se viver, você tem uma meta clara, que é aumentar o IDH, você tem uma contribuição muito clara da saúde, que é reduzir a mortalidade infantil, e você tem um projeto cujo objetivo é fazer essa redução. Então é um programa que tem uma linha de ação, de atuação muito grande, muito clara, um inter-relacionamento de planejamento muito forte, além do apelo para a sociedade, da motivação política para o governo de trabalhar em prol da redução da mortalidade infantil e todos os aspectos que envolvem, no ponto de vista da organização do sistema de saúde, que se requerem. Então, é uma meta mobilizadora, um norte muito forte. Então eu acho que isso tudo contribuiu para a escolha do Viva Vida. (EQUIPE DO PROJETO VV2)

Para outros quatro entrevistados, o Banco Mundial, como patrocinador da empreitada de avaliações e reconhecedor do destaque do projeto Viva Viva, quase que *exigiu* que o projeto estivesse entre o rol de políticas avaliadas. Para o entrevistado VV1, estava *na moda* monitorar o projeto Viva Vida.

> Na época, quando veio essa ideia de Centro do Viva Vida, como era um diferencial do Estado de Minas deve ter feito muita propaganda disso para o Banco Mundial "Bacana, vamos fazer um negócio inovador...", então eu acho que tudo conspirou para que isso fosse o objeto principal da avaliação. Eu acho que nesse relacionamento SES, Estado para Resultados e Banco eu receio que não tenha sido exatamente assim, uma demanda Saúde, que tenha sido mais uma coisa "Bacana, Saúde, você veio, vendeu para a gente essa ideia inovadora, só que o Banco trabalha da seguinte forma...". Acho que o Banco tem talvez o papel mais de protagonista mesmo nesse processo e era também uma proposta do Estado para Resultados ter lá um setor, um lócus de avaliação de políticas. (GERENTE DO PROJETO VV, grifo da autora da dissertação)

> No programa Viva Vida era uma exigência, não sei se posso falar 'exigência', mas o Banco Mundial esperava que esse programa fosse avaliado. Então eu acho que foi esse clima, esse clima aí que vigorou o Estado nessa época: programas acontecendo, interesse de avaliação. (AVALIADOR VV2, grifo da autora da dissertação)

> Porque ele era estruturador, porque ele tinha, (não lembro se era exatamente por isso) mas tinha uma questão do financiamento do Banco Mundial que envolvia isso, é um volume expressivo de dinheiro, era uma iniciativa inovadora também. Então eu acho que na época estava na moda monitorar ele. (EQUIPE DO PROJETO VV1)

Nota-se, em um primeiro momento, que os motivos para a avaliação do projeto Viva Vida estavam mais claros, na percepção da equipe do projeto e dos avaliadores, que em outros projetos avaliados. No entanto, mesmo existindo motivos mais claros para a concepção da avaliação do projeto *carro chefe* da SES, isso não quer dizer que a avaliação tenha sido bem recebida pelos técnicos e gestores da secretaria. É possível observar, pelo relato do entrevistado do EpR e do gestor do projeto, que o projeto Viva Vida não foi a primeira opção

da Secretaria de Saúde. A ideia inicial proposta da SES era diferente da proposta da coordenação do projeto de avaliações, proposta esta que não foi aceita pelo EpR e pela Fundação. Para o gestor do projeto, a proposta inicial, e coerente, da SES seria avaliar o impacto das redes de saúde como um todo, já que o desenho dos projetos na saúde é interdependente aos três níveis de atenção, como relatado pelos entrevistados.

> Eu sei que nesse processo de definição, a secretaria queria avaliar o sistema de saúde, a rede de atenção e não só o Viva Vida, e aí, por algum motivo, acho que questões técnicas, própria discussão com o Banco Mundial com direção do EpR, definiram que não dava para avaliar, tinha que escolher um dos pontos dessa rede. E aí optou-se pelo Viva Vida, que era um projeto que estava num estágio mais inicial, então a gente já tinha alguns centros já implantados. (TÉCNICO EPR1, grifo da autora da dissertação)

> A insistência em avaliar a rede como um todo, ela era muito pertinente, porque realmente não era o centro que ia transformar a mortalidade infantil na região, ele [o Centro Viva Vida] é mais um ponto para colaborar com isso e tal. Então nesse processo do planejamento teve outro fator, como não era exatamente uma encomenda nossa, em algum momento o Estado para Resultados junto com a Fundação lá... (GERENTE DO PROJETO VV)

O desenho da avaliação foi, portanto, formatado pela FJP e pelo EpR, com foco no projeto Viva Vida. O contato com a avaliação já com seu desenho definido, por parte dos integrantes da equipe do projeto, foi *institucional* de acordo com o entrevistado VV2. Para o entrevistado VV1, a demanda foi *top down,* e a equipe do projeto foi avisada que uma avaliação de impacto iria ocorrer, sem maiores justificativas de sua importância e como a equipe do projeto poderia se engajar no processo, como também confirma o avaliador VV2 ao reconstituir o processo de definição da avaliação.

> E aí a gente ficou sabendo da avaliação institucionalmente falando, só ligou, se apresentou, explicou qual que era o motivo, para que que havia sido contratado, e a gente fez as entrevistas e colaborou com o que o pessoal que pediu de informação. (EQUIPE DO PROJETO VV2, grifo da autora da dissertação)

> O que eu lembro é que essa demanda não partiu da coordenação porque ela veio de cima para baixo. Eu me lembro de a gente sendo avisado, mas não de a gente sendo incluído nesse processo de "vamos encomendar uma avaliação para que que serve..". Quando eu fiquei sabendo da avaliação já estava definido que ela ia acontecer. (EQUIPE DO PROJETO VV1)

> Eles chamaram especialistas, consultores, para discutir com eles assim "Os programas, o que vocês acha? Vocês acham que eles podem ser avaliados?". Eles fizeram um documento inicial que era o estudo de viabilidade da avaliação, se a avaliação tinha condição de ocorrer. E eles, então, do ponto de vista deles eles chegaram à conclusão que poderia ser avaliado. Agora, o pessoal da secretaria, realmente eu tenho quase que certeza absoluta, eles foram chamados para explicar,

> mas, assim, eles já foram comunicados que esses dois programas tinham sido selecionados e houve essa discussão, que eles achavam que principalmente o Viva Vida ainda não cabia uma avaliação de impacto. (AVALIADOR VV2)

De acordo com Patton (2008), uma avaliação é capaz de evocar fortes emoções, associações negativas e um medo genuíno aos *stakeholders* que não estavam envolvidos com a avaliação e desconhecem o seu propósito original. Para o autor, o estabelecimento de uma fundação sólida com para o acontecimento de qualquer avaliação, com entendimentos compartilhados e propósito claro, é ainda mais importante que o desenho de métodos sofisticados. Na avaliação do projeto Viva Vida, essa *fundação* não foi criada. O entendimento mútuo entre os atores envolvidos não foi criado no início da avaliação, e muitas percepções negativas foram criadas à época. As expectativas iniciais por parte da SES já começaram com uma visão perversa das implicações da avaliação, antes mesmo de dar início à avaliação, os entrevistados já imaginavam que geraria um impacto negativo, como comentam o gerente do projeto e os integrante da equipe.

> A gente sabia que essas políticas novas na área de saúde – a implantação de novos pontos de atenção e a própria questão da implantação das redes, elas não geram impacto imediato porque elas dependem também de mudança de cultura, ajustes graduais de processo etc. A gente tinha essa preocupação também. Já pensou a gente expor um final de avaliação num fórum, que é o controle social, um resultado negativo? Acho que essa perspectiva do avaliado ter um pouquinho de temor, ela perpassou o processo o tempo todo. E foi nesse contexto que foi feito o planejamento da avaliação. (GERENTE DO PROJETOVV, grifo da autora da dissertação)

> Ninguém achava que [a avaliação] seria a salvação. Porque a gente tinha tanto problema. Porque a gente tinha muito problema operacional. O Centro estava engatinhando, tinha um monte inaugurando ao mesmo tempo. Ela veio, eu acho, que talvez no momento inoportuno porque era um momento em que a gente não ia poder parar muito para ver. (EQUIPE DO PROJETO VV1, grifo da autora da dissertação)

> A nossa expectativa era de que a complexidade do objeto do programa mais as ações que foram feitas eventualmente ao longo do tempo, essas ações pontuais que não tiveram uma reincidência, uma recorrência para frente, que isso fosse nublar um pouco os aspectos positivos do programa do ponto de vista do impacto dele. Eu estou te dizendo é que em termos dos resultados a gente já esperava que a nossa avaliação de impacto não seria muito positiva, não seria o melhor. Tipo assim, "nós entramos com o programa e de repente a mortalidade desapareceu". Não! A gente já sabia que não. (EQUIPE DO PROJETO VV2, grifo da autora da dissertação)

Não é exagerado pensar que a avaliação de políticas, muitas vezes, é percebida nos setores sociais como um *monstro*, como destaca Mokate (2002). A avaliação como um *monstro*, cunhada por Mokate (2002), ocorre quando a avaliação aparece como uma tarefa a ser feita, geralmente porque foi imposta por uma exigência legislativa, como uma obrigação colocada

pelo governo central ou até como reivindicação de agências de fomento, como é o caso do Banco Mundial. Essa característica impositiva, para a autora, é um dos motivos responsável pelas percepções negativas quanto ao processo de avaliação e que tendem a refletir uma não utilização dos estudos, sem uma maior contribuição para o desenho ou gestão da iniciativa avaliada. No caso da avaliação do projeto Viva Vida, é possível observar que a equipe do projeto, bem como o gerente, assumiu a avaliação como uma obrigação que, para os entrevistados não estaria atrelada às necessidade dos projetos à época. Assim, a avaliação foi percebida como uma intervenção externa ao projeto e que se converteu, rapidamente, em um processo pouco aliado à equipe gestora do Viva Viva. Na visão dos avaliadores, a resistência da SES quanto à avaliação do projeto Viva Vida era perceptível. Para o avaliador VV2, era aparente que a equipe do projeto não acreditava que a avaliação de impacto seria uma opção que auxiliaria o projeto naquele momento, já que nem um monitoramento consistente existia. O avaliador VV1 concorda que a resistência permanecia, mas que, como avaliadores contratados pelo EpR, eles não poderiam deixar de fazer a avaliação.

> O programa estava ainda muito no início. Eu lembro de que uma das pessoas presentes ainda foi categórico, ele falou: "Esse programa precisa de um programa de monitoramento consistente. Está na hora ainda de a gente pensar de um monitoramento bem sério para o programa, ainda não é hora da avaliação". (AVALIADOR VV2)

> Como que a gente decidiu o que avaliar; como avaliar? Primeiro a gente foi conhecer um pouco do programa. No início a gente viu que não tinha assim uma receptividade tão grande por parte da Secretaria de Saúde, porque tem sempre aquela coisa de ser avaliado não é algo muito bom, então a gente sentiu que existia um pouco de resistência, principalmente porque a Fundação sempre teve uma parceria muito forte com a Secretaria de Saúde. Mas agora a gente não estava sendo contratado pela Secretaria de Saúde, como um órgão externo para avaliar a Secretaria de Saúde, a gente estava sendo contratado pelo EpR. (AVALIADOR VV1, grifo da autora da dissertação)

Além dos avaliadores da FJP, os próprios coordenadores do processo de avaliação, localizados na FJP e no EpR, sabiam da resistência da SES de que o projeto Viva Vida não estava preparado para passar por uma avaliação de impacto. Para o técnico do EpR, a equipe do projeto assumiu a avaliação quase como uma *burocracia*, mais para *cumprir tabela*, pouco preocupados com os resultados futuros da avaliação.

> Na saúde não havia muito entusiasmo com essa avaliação também. A percepção lá [na secretaria de saúde] é de que era prematuro fazer a avaliação do programa, que eles não tinham tido tempo de amadurecer. Mas isso não foi um obstáculo para que se obtivesse os dados e a colaboração deles. (COORDENAÇÃO FJP)

> Resistência no sentido de utilizar os resultados da avaliação depois. Assumiram quase que como uma burocracia "Tem que fazer, vamos fazer". Mas não se furtaram a fornecer informações nem nada não, mas não se envolveram. Acho que eles fizeram a apresentação do que que era o Viva Vida e tal mais para cumprir tabela. Pouco preocupados com os resultados que iriam sair dali. Eu acho que eles ficaram meio que como, não é marido traído não, mas assim, não assumiram aquilo como... Não era uma demanda legitima deles essa avaliação. (TÉCNICO EPR1, grifos da autora da dissertação)

O contexto inicial da avaliação do projeto Viva Viva foi marcado pela resistência da equipe do projeto, tanto em relação ao objeto da avaliação, já que na visão da SES seria mais apropriado avaliar a rede de saúde como um todo, quanto à maneira com que a avaliação foi transmitida, como um comunicado institucional, sem dar chances à equipe do projeto de negociar. Em função desse contexto, os envolvidos na Secretaria de Saúde pouco se preocuparam com os resultados que poderiam ser evidenciados a partir da avaliação, antecipando possíveis conclusões negativas quanto ao impacto do projeto. Assim, assumiram a avaliação como um processo imposto por indivíduos do alto escalão do governo, com um nível maior de autoridade, e se limitaram a uma atitude mecânica quanto ao processo de avaliação, como reconstruído na próxima seção a seguir que trata do processo de implementação da avaliação.

7.2 Implementação do processo de avaliação e envolvimento dos atores

Com uma concepção tumultuada, a avaliação do projeto Viva Vida contou com participações pontuais da Secretaria de Saúde que, naquele momento, tinha outras preocupações quanto ao desenvolvimento das ações do projeto. Tanto na visão do gerente do projeto, quanto do corpo técnico do EpR e do avaliador da FJP, o processo de planejamento e implementação da avaliação teve um baixo envolvimento da SES, cuja participação se limitou ao fornecimento de dados e informações, pouco interessada nos resultados da avaliação. Quem liderou o planejamento e implementação da avaliação, na percepção do gestor, foi a FJP e o EpR, que organizaram o cronograma de avaliação de acordo com as metas vislumbradas pelo Banco Mundial, se atendo mais aos prazos e à necessidade do contrato com a instituição de fomento que às prioridade da SES.

> É. Assim. Planejamento mesmo, o que seria avaliado, como que seria feito e tal, envolvimentos pontuais em reuniões, com um desenho maior da Fundação e o próprio Estado para Resultados. Execução, de novo, muito assim "gente, nós estamos precisando aí da lista de gestantes numa micro". Então, de novo, a SES muito como apoio da execução. Em momento nenhum a gente debruçou ativamente

> sobre o processo e tal, ficou muito a perspectiva de uma avaliação externa mesmo com o desenho externo também. (GERENTE DO PROJETO VV)

> Bom, a parte de definição de escopo, pensando em cronograma e recursos necessários, tudo foi discutido entre o Estado para Resultados e a Fundação com pouca participação da secretaria. E muito em função daquela meta final que tinha porque eles tinham que cumprir o indicador lá com o Banco Mundial. Então quase que um negócio de trás para frente. (TÉCNICO EPR1)

> Assim, a parte inicial eles participaram. Assim, para a avaliação mesmo eles não participaram da avaliação não, porque a gente estava avaliando um programa deles, então existe certa distância. Mas eles participaram fornecendo dados e algumas informações. Por exemplo, informações de repasses, informações de números de Centros Viva Vida, localização desses centros. No segundo momento também informações sobre a gestante de alto risco, que a gente utilizou para avaliação de impacto. Então, eles forneciam algumas informações disso. Informações sobre como que era o programa, funcionamento do programa agente tinha reuniões com eles para entender o programa, as ações. (AVALIADOR VV1)

Na visão da equipe do projeto Viva Vida, a SES não foi alijada do processo de avaliação, o momento em que a avaliação ocorreu, simplesmente, não era oportuno, como comentado em entrevistas. Apesar de os avaliadores da FJP insistirem na participação da SES e os convidarem a participar da avaliação, como afirmado pelo entrevistado VV2, as prioridades da equipe do projeto eram outras. Como completa o entrevistado VV1, na época, a equipe estava mais focada em inaugurar novos Centros Viva Vida do que rever os processos e verificar se o projeto gerava resultados. As forças da equipe estavam concentradas nos serviços e produtos elencados no Acordo de Resultados que geravam uma pressão por entregas. A avaliação seria uma atividade *paralela, tocada por outra instituição* que não era prioridade para o Acordo de Resultados.

> A gente ficava direto com a Fundação, e eu não posso falar que a gente foi alijado do processo não, de jeito nenhum. Eu sinto que nesse processo avaliativo a gente foi bastante participativo. Eu não quero dizer que a gente contribuiu muito, é que a gente foi... chamado a participar... convidado, é nesse sentido. Assim, houve uma abertura muito grande por parte de quem estava avaliando, houve uma receptividade de lá em nos fazer sentir elemento importante da avaliação. Era muito claro assim, "sem vocês, sem as informações que vocês têm não há avaliação, não tem como se efetivar o que a gente precisa". E isso ficou claro desde o início. (EQUIPE DO PROJETO VV2)

> Era uma época que tinha muita inauguração, tinha uma pressão por isso. A gente estava muito mais focado em inaugurar novos Centros do que rever mesmo o processo, entender um pouco melhor da iniciativa. Ela veio no período que foi muito conturbado nesse sentido. Tinha uma série de outras atividades que tinham por força do acordo de resultados que ia acontecer, então, para a gente ela era uma iniciativa paralela, tocada por outra instituição, que podia contribuir com alguma coisa, mas que naquele momento não era a nossa prioridade, mesmo porque nem era uma atribuição nossa, a gente estava mais como facilitador, mais para abrir portas para o

> pessoal. Aqui [na Secretaria de Saúde] tudo é para ontem, você é atropelado, você não tem tempo de parar e rever. O varejo das decisões que a gente tem que tomar, da tramitação da administração aqui dentro, da condução do projeto, não permite que a gente tenha dedicação. Eu gosto muito disso, ainda me envolvo com a universidade, então sei que isso demanda tempo. O argumento é físico, não dá para estar em dois lugares ao mesmo tempo, então a gente não consegue. Por mais que a gente queira, por mais que a gente precise disso até para melhorar internamente, do jeito que a própria secretaria é organizada, ela não favorece que isso aconteça. (EQUIPE DO PROJETO VV1, grifo da autora da dissertação)

De acordo com Mokate (2002), a gestão dos setores sociais é tradicionalmente caracterizado pelo foco em atividades: um estilo de trabalho orientado para o desempenho de planos de trabalho predefinidos, em que as atividades se tornam fins em si mesmas. A equipe de trabalho é orientada por atividades de conformidade e processos, em que a máxima desse estilo de trabalho é a entrega de um bem ou serviço, o que requer apenas informações de monitoramento de processos. Dessa forma, não existem interesse e incentivos para realizar ou para solicitar uma avaliação dos resultados ou dos impactos de um determinado projeto. Como aconteceu no projeto Viva Vida, as equipes de trabalhos não questionam o objetivo final de seus projetos e não verificam seu cumprimento porque não foram responsabilizados pelo alcance desses objetivos, mas para o desempenho das atividades propostas que, no caso em questão, estariam elencadas no instrumento de contratualização do acordo de resultados.

O papel do EpR, de acordo com a percepção do técnico envolvido na avaliação, era fazer a mediação entre a Secretaria de Saúde, a FJP e o Banco Mundial, como afirmado no primeiro trecho de entrevista selecionado abaixo. Entretanto, na visão dos *stakeholders* da avaliação, a visão quanto ao papel do EpR muda de conceito, dependendo do ator que o julga, se é avaliador ou se trabalha na equipe do projeto avaliado. Se o ator constitui a equipe do projeto avaliado, como é o caso do segundo depoimento abaixo, o EpR é visto como um obstáculo que atrapalhava o desenvolvimento natural das ações do projeto. Nas palavras do entrevistado VV1, "ele é mais um que vem ocupar o seu tempo". Por outro lado, na visão do avaliador VV1, o EpR é percebido como um parceiro técnico, responsável por fornecer os dados necessários à avaliação e por dar diretrizes profissionais ao grupo de avaliadores. Assim, como demonstram os trechos abaixo, o papel do EpR vai de *vilão* à *mocinho*, dependendo da percepção do *stakeholder* que o avalia.

> O papel do Estado para Resultados era fazer essa relação entre o Banco Mundial, que prestava uma consultoria nessa avaliação, e a Fundação João pinheiro, e a secretaria, que era a parte de fornecimento de dados, o acordo da secretaria. Então, o papel do Estado para Resultados era mais esse, de fazer as relações. A gente não participou de definição metodológica, estimular essa discussão entre esses entes e,

> junto à secretaria a gente fazia a relação da Fundação com a secretaria para conseguir os dados que fossem necessários para a avaliação. (TÉCNICO EPR1)

> Eu acho que ele [Programa Estado para Resultados] existe porque tem o interesse dele existir, eu não acho que ele agrega. Alguma coisa certamente deve agregar sim, mas na rotina do programa ele é mais um que vem ocupar o seu tempo. Ele não é alguém que fala "Ah, que bom! É o salvador!". Não! É tipo: "Que saco, mais uma reunião!". É isso que acontece, mas é porque é um problema da secretaria, de não ter equipe completa, de ter poucas pessoas fazendo muitas coisas ao mesmo tempo. (EQUIPE DO PROJETO VV1)

> Eu acho que o escritório também foi um grande parceiro nosso. Eles tinham interesse de entender, de ter os resultados. Em momento algum o escritório pressionou para a gente mostrar que o programa era lindo, maravilhoso, que tinha impacto. Não! Eles ajudaram sempre, o tempo todo fornecendo as informações, conseguindo os contatos, conseguindo listas de dados. Às vezes, as informações demoravam e a gente ligava e eles pediam, tentavam agilizar para que as coisas acontecessem de uma forma mais rápida. (AVALIADOR VV1)

Para a implementação da avaliação, o primeiro passo dos avaliadores foi procurar entender o projeto como um todo para, assim, poder aprofundar o desenho da avaliação. No entanto, de acordo com a percepção do avaliador VV1, explorar o projeto e selecionar a abordagem avaliatória, não foi algo trivial para a equipe de avaliadores da FJP.

> Então, entender o programa já não era algo tão fácil, e a gente demorou um pouco para entender todas as ações, entender como os recursos foram distribuídos, quais os municípios foram eleitos, quem participaria do programa, então, isso tudo levou um tempo para a equipe. E a partir disso, a partir de entender o que era o programa, a gente foi ver o que a gente iria fazer em termos de avaliação. A gente pensou em fazer a avaliação de impacto de cara nessa primeira avaliação, mas... O mais difícil foi decidir o que fazer, como fazer. (AVALIADOR VV1, grifo da autora da dissertação)

De acordo com o avaliador VV1, o projeto era muito grande e não haveria condições de ser avaliado abrangendo todas as suas ações; não haveria nem tempo hábil, nem recursos financeiros suficientes, como mostrado no relato abaixo. A estratégia dos avaliadores foi então escolher a ação de implantação dos Centros Viva Vida para a avaliação. Propôs-se avaliar o impacto do CVV sobre as condições de saúde e utilização dos serviços de saúde das gestantes de alto risco e crianças nascidas dessa gestação. Ao contrario de medir a diminuição direta da mortalidade materna e da infantil em Minas Gerais, optou-se por verificar os efeitos do atendimento de pré-natal e parto sobre a morbidade materno-infantil, considerada evitável ao permitir o diagnóstico e tratamento de doenças acometidas durante a gravidez. A justificativa para tal escolha está na quantidade de recursos dispendidas por esta ação – que compreende a construção dos centros nas microrregiões do estado, a compra e instalação de

equipamentos, a alocação e treinamento dos profissionais de saúde, etc – além de constituir a principal ação do projeto, que para o entrevistado era *a menina dos olhos.*

> Um programa muito grande e a gente não teria como avaliar o programa como um todo; a gente não teria tempo hábil, a gente não teria nem recursos financeiros para fazer uma avaliação do programa como um todo, então a gente tinha que escolher alguma ação. E aí a gente centrou no Centro Viva Vida. Porque o Centro Viva Vida? Porque o Centro Viva Vida sempre foi considerado a principal ação, a *menina dos olhos* do programa, e, também a ação que recebeu maior número de recurso. (AVALIADOR VV1)

Na tentativa de amenizar o *desconforto* da equipe do projeto com a avaliação de impacto do Viva Vida e, também, de respeitar o estágio em que o projeto se encontrava, a equipe de avaliadores da FJP sugeriu que, além de uma avaliação de impacto do Centro Viva Vida, se realizasse primeiro uma avaliação dos processos do projeto. Assim, anteriormente à mensuração dos impactos, seria realizada uma avaliação da implementação geral do projeto, na como relatado na fala do avaliador VV2.

> Claro que uma avaliação externa, ela sempre vai causar algum desconforto. Mas, de qualquer forma, eu também entendi que tinha essa coisa de amadurecimento. A gente até propôs: "por que a gente não pensa numa avaliação, numa avaliação da implementação e não apenas de impacto?" E isso foi aceito. Então, a gente fez, na verdade, uma avaliação do processo de implementação e da avaliação de impacto. (AVALIADOR VV2)

O desenho da pesquisa consistiu, portanto, em avaliar o processo de implementação do projeto Viva Vida e, posteriormente, o impacto do Centro Viva Vida. A metodologia de avaliação proposta procurou considerar as principais características do projeto e suas especificidades, bem como o estágio em que o projeto se encontrava desde a sua implementação. Na análise de processo, o objetivo foi verificar se os recursos alocados foram utilizados como previsto pelo projeto Viva Vida. Assim, como resumido pelo avaliador VV1, a avaliação do projeto Viva Vida incidiu em dois tipos distintos de estudos, processos e impacto.

> Foram duas avaliações, dois anos diferentes, no primeiro ano a gente optou por fazer avaliação de processo. Por que avaliação de processo e não avaliação de impacto? Porque a gente sentiu que a gente estava precisando entender mais o programa e que não era, talvez, o momento ideal para fazer uma avaliação de impacto. O programa não estava maduro suficiente para a gente fazer avaliação de impacto, então, fez-se poção por uma avaliação de processo. (AVALIADOR VV1)

A avaliação de processo permite detectar falhas na implementação do projeto que podem se refletir no resultado final esperado (AGUILAR; ANDER-EGG, 1995). De acordo com o relatório final de avaliação do projeto, o objetivo da avalição de processos seria verificar se os recursos alocados estavam sendo utilizados como previsto, se as atividades planejadas estavam sendo realizadas conforme estabelecido inicialmente e se as instituições beneficiadas contavam com os equipamentos e *kits* de acordo com o planejado (FUNDAÇÃO JOÃO PINHEIRO, 2010a). Em síntese, o estudo foi efetivado com base em entrevistas em instituições sorteadas que receberam recursos do projeto, quais sejam, unidade básica de saúde, maternidade de risco habitual, UTI neonatal, Centro Viva Vida e Secretaria Municipal de Saúde. O questionário administrado nos CVV, além de avaliar a presença, instalações e condições de uso dos equipamentos fornecidos pelos *kits* e para os CVV, foi também investigado o processo administrativo das unidades beneficiadas pelo PVV e sua interligação com as secretariais estaduais e municipais de saúde no que se refere ao monitoramento das metas e da produção (FUNDAÇÃO JOÃO PINHEIRO, 2010a).

A segunda fase da avaliação, de mensuração do impacto, foi realizada considerando resultados esperados pelos CVV no atendimento às gestantes de alto risco e às crianças nascidas dessa gestação (FUNDAÇÃO JOÃO PINHEIRO, 2010a). Foi utilizado o método de pareamento por escore de propensão[6] considerando dez diferentes indicadores de saúde impactados pelo projeto Viva Vida. Foram selecionados 37 municípios pelo método de pareamento por escore de propensão, além de todos os municípios-polo de sua respectiva microrregião de saúde. Foram consideradas na análise apenas as gestantes de alto risco que tiveram, pelo menos, um contato com o SUS (FUNDAÇÃO JOÃO PINHEIRO, 2010a).

Quanto à seleção dos métodos, no entanto, alguns atores (YEH, 1980) apontam que métodos menos sofisticados proporcionam um uso melhor das avaliações como, por exemplo, Leviton e Boruch (1983), que argumentam que estudos avaliatórios focados em desenho ou processos

---

[6] A avaliação de um programa ou intervenção por meio da metodologia do pareamento pelo escore de propensão tem por objetivo comparar participantes (grupo tratamento) e não participantes (grupo controle) do programa, tomando como base para a comparação similaridades em termos de características observáveis. O método, desenvolvido em Rosenbaum e Rubin (1983) é útil quando se considera que apenas as variáveis observáveis afetam a probabilidade de um indivíduo participar do programa, permitindo uma focalização mais precisa e também uma alta taxa de adesão dos beneficiários à intervenção. De acordo com Khandker, Koolwal e Samad (2010), o método do pareamento baseado no escore de propensão baseia-se na construção de um grupo de comparação estatística fundamentado em um modelo de probabilidade de participação, utilizando características observáveis. O efeito médio de ser beneficiado pelo programa é calculado a partir da diferença média entre os resultados desses dois grupos.

seriam mais bem utilizados que estudos focados em resultados dos programas. Esse argumento é compartilhado por um dos avaliadores do projeto que acredita que, do ponto de vista do usuário, as avaliações de processos seriam mais úteis que os tipos que utilizam métodos quantitativos sofisticados, dado que se concentram em problemas do dia-a-dia do projeto.

> Agora, se você me perguntar "qual a avaliação que você achou mais interessante?". Qual que você acha que é a avaliação que foi melhor em termos de resultados mais concretos para a secretaria, eu acho que foi a primeira avaliação; foi a avaliação de processo. Eu acho que para a secretaria a avaliação de processo foi mais interessante. Eu gostei muito da avaliação de processo, eu acho que a avaliação de processo, ela é muito interessante porque, além dela detectar o que está acontecendo, as principais dificuldades, eu acho que ela consegue dar uma informação maior: "Os principais problemas, os principais pontos fracos estão aqui. Porque que estão aqui? O que que está acontecendo?". (AVALIADOR VV1, grifo da autora da dissertação)

As dificuldades atreladas às avaliações não incidem somente na definição da metodologia, mas também em questões operacionais, como, por exemplo, problemas na obtenção de informações sobre os programas avaliados (ARRETCHE, 1998). Um ponto comum na percepção dos diferentes envolvidos na avaliação do Viva Vida diz respeito à dificuldade em conseguir os dados para a pesquisa. O processo de coleta de dados, principalmente para a avaliação dos impactos do Centro Viva Vida, não foi uma etapa fácil do estudo, de acordo com avaliadores, com o próprio gerente do projeto e com o técnico do EpR.

> Para a gente não foi algo muito fácil, porque a gente sabe que existe essa distribuição, mas, por outro lado, a gente tem toda aquela dificuldade de obter o dado. Como foi essa distribuição, a gente tem sempre relatos, mas, muitas vezes, as informações não estavam consolidadas, não estavam fáceis para a gente obter. (AVALIADOR VV1)

> Uma das dificuldades da avaliação do Viva Vida foi com relação a dados, porque o registro... A gente pediu para as Gerências Regionais de Saúde que nos encaminhasse a relação de gestantes daqueles municípios, dos municípios ou das microrregiões, e isso era obtido junto às unidades básicas de saúde, o PSF e tal, só que esse registro é precário. Então, assim, muita coisa a gente não recebeu e muita coisa vinha assim, tipo uma folha escrita à mão e a gente tinha que processar isso aqui. A gente, "eu". (TÉCNICO EPR1)

> Mas na época eu lembro que foi um pouco de sofrimento. A gente tinha que jogar isso para a micro para tentar conseguir com o município a lista de gestantes para o pessoal correr atrás. Não foi um processo fácil. A participação da secretaria foi mesmo quando demandada tentar dar os caminhos para providenciar isso e aí eu sei que a pesquisa foi executada incluindo entrevistas com as gestantes, as unidades lá da ponta. Aproveitaram e fizeram um diagnostico também das estruturas locais. Eu me lembro deles trazerem esses dados. (GERENTE DO PROJETO VV, grifo da autora da dissertação)

Lobo (2001) reforça a importância da informação desde o início de qualquer programa ou projeto governamental. Para a autora, sistemas de informação bem estruturados possibilitam estratégias de avaliação mais condizentes com seus objetivos. A dificuldade de obtenção dos dados foi, na visão de diferentes *stakeholders*, a comprovação de que o projeto não estava preparado para ser avaliado. Na verdade, como ressalta o avaliador VV2, os gestores programas governamentais deveriam priorizar as avaliações no momento em que são desenhados, já delineando as suas necessidades de organização de dados.

> Não considero que o programa estava pronto para ser avaliado porque não tinha linha de base, não tinha nada. O projeto não era executado pensando em algum momento ter uma avaliação, então se fosse feito pensando nisso o próprio registro da gestante seria algo mais bem estruturado. (TÉCNICO EPR1, grifo da autora da dissertação)

> E aí eu vejo nitidamente que a gente não tem essa preocupação de gerar uma linha de base para fazer uma avaliação. Então, assim, nasce a política, depois você tem uma ideia de avaliação e você tem que correr atrás. Olha só como que as coisas são meio atropeladas! (GERENTE DO PROJETO VV, grifo da autora da dissertação)

> Eles [equipe do projeto Viva Vida] tinham um problema até na lista dos beneficiários. Quer dizer, conseguir uma lista de beneficiários não foi uma coisa trivial, para você ver como que o programa não estava nem organizado ainda em termo de implementação. Uma coisa muito importante também, por exemplo, que as políticas implementem, incorporem essa questão da avaliação no momento em que a política está sendo formulada. Isso é uma questão crucial. Por que a gente encontrou dificuldades às vezes... Quais indicadores que devam ser avaliados? Isso aí é porque no momento em que foi pensando o trabalho não se pensou na avaliação, porque quando se pensa você já formula coisas assim, de tal forma, objetivos bem específicos que facilitam a avaliação. Você estabelece fluxos de informação que vão te possibilitar fazer uma avaliação melhor, quer dizer, você prepara o terreno para a avaliação. Na verdade, seria importante que isso fosse incorporado no Estado, que a avaliação, ela fosse pensada já no momento do programa. Ou seja, ao formular um programa eu deveria pensar "é importante que esse programa seja avaliado no dia", então eu já posso preparar o terreno, garantir que as informações vão sendo sistematizadas, pensando em objetivos bem claros, porque, às vezes, você encontra objetivos ambíguos no programa, então a hora que você vai avaliar... Enfim, acho que isso é importante, se pensar na avaliação no momento do desenho do programa. (AVALIADOR VV2, grifo da autora da dissertação)

Observa-se, assim, um processo de implementação marcado pelo distanciamento da SES quanto ao planejamento e implementação, limitando-se ao fornecimento de dados regulado pelo EpR. Quanto à execução do estudo avaliatório, notam-se dificuldades operacionais ao longo da coleta de dados o que, na visão de alguns atores, demonstrava que o projeto não estava pronto para ser avaliado. Apresentados a divisão de papel entre os atores envolvidos na avaliação e o desenvolvimento metodológico da avaliação, a seção seguinte remonta o processo de finalização da avaliação do Viva Vida, o que inclui, tanto seus resultados, quanto

a forma como foi encaminhada para a SES e a disseminação dos resultados dentro da secretaria.

7.3 Resultados e processamento de informações

Ao final de 2009, foi concluída a primeira fase da avaliação, que compreende uma avaliação do processo de implantação do programa. A segunda fase, que corresponde à avaliação de impacto dos Centros Viva Vida, foi concluída em dezembro de 2010. Em linhas gerais, como resultados principais da avaliação de processos destacam-se fragilidades relacionadas a recursos humanos, como carência de profissionais voltados para o atendimento da mulher e da criança, alta rotatividade, falta de comprometimento com o programa e necessidade de capacitação; problemas na interação entre atenção primária à saúde e atenção secundária, sendo essa constatação particularmente preocupante uma vez que as UBS (atenção básica) têm papel fundamental no encaminhamento das gestantes de alto risco aos CVV (atenção secundária); e, por fim, ausência de estratégias de divulgação do projeto, em especial, dos CVVs, junto à população e profissionais do setor de atenção primária principalmente nos demais municípios que compõem a microrregião de saúde (FUNDAÇÃO JOÃO PINHEIRO, 2010a).

Na avaliação de impacto dos CVVs, concluída após um ano da avalição de processos, os resultados sugerem que o CVV causou impacto apenas em um único indicador de dez indicadores de saúde considerados[7] (FUNDAÇÃO JOÃO PINHEIRO, 2010a). Além da estimação do impacto do Centro Viva Vida, foi analisada a focalização dos mesmos, já que a principal dificuldade encontrada se refere aos encaminhamentos realizados pelo setor de atenção primaria aos CVV. De forma simplificada, os problemas de focalização identificados estavam relacionados com o deslocamento das gestantes de alto risco para os municípios-sede do CVV e, portanto, coma carência de transporte e com a comunicação do projeto a atores-chave. O avaliador VV2 aponta que os resultados da avaliação somente confirmaram uma perspectiva inicial apresentada pela SES: de que o projeto não apontaria impacto.

---

[7] O único indicador que apresentou aumento foi o número de partos pré-termo (partos em que ocorre nascimento prematuro do recém-nascido, menos de 37 semanas de idade gestacional). O fato do CVV ter aumentado o número de partos pré-termo sugere, a princípio, um impacto até inverso ao esperado pela implantação do projeto. No entanto, como afirmado no relatório de avaliação, seria necessário maior aprofundamento em relação ao acorrido, já que, no caso de gestação de alto risco, partos prematuros podem ser considerados benéficos para a saúde da gestante de alto risco e da criança nascida dessa gestação (FUNDAÇÃO JOÃO PINHEIRO, 2010a).

> O que eu vejo do Programa Viva Vida é o seguinte: não houve impacto. As conclusões, tá? Mas é aquilo que a gente, de alguma forma, estava desconfiada, os próprios dirigentes da SES falaram: "Olha, está muito cedo ainda para se falar em impacto". O programa estava começando e estava enfrentando dificuldades básicas. (AVALIADOR VV2)

Apesar de o projeto não apresentar impacto quanto à atuação do CVV, de acordo com o resultado da avaliação, os avaliadores fazem ressalvas quanto às implicações dessa conclusão. Para o avaliador VV, não houve impacto nos indicadores que foram selecionados para a avaliação, mas não seria possível afirmar, com 100% de certeza, que o projeto teria que ser interrompido devido a esse resultado. Assumir mudanças mais efetivas no projeto somente com base nas informações derivadas da avaliação seria muito arriscado e, por isso, foi necessário advertências para a equipe do projeto.

> A gente sempre deixou bem claro que não teve impacto nos indicadores selecionados dentro daquilo que a gente selecionou para avaliar: a gente sabia das limitações dessa avaliação, a gente sabia da questão que o programa precisava de um tempo maior de maturação para ser avaliado. Então, a gente deixou bem claro isso, que não teve impacto, mas não é que o programa... "Ah, mas então o programa tem que ser descartado?". Isso foi até afirmado: "Não teve impacto, o programa tem que ser descartado?". Não! A gente não poderia assumir isso. A gente fez assim, várias ressalvas nesse sentido: "Dentro daquilo que a gente avaliou, dentro das limitações dessa avaliação, dos indicadores selecionados a gente viu que teve impacto apenas em um indicador, mas não quer dizer que o programa não esteja no caminho certo". Então a gente reforçou bem esses resultados. (AVALIADOR VV1, grifo da autora da dissertação)

Quanto às ressalvas feitas pelo avaliador VV1, torna-se importante ressaltar a variável credibilidade apresentada por Leviton e Hughes (1981). Para os autores, os administradores utilizam as avaliações também para aumentar a certeza sobre conhecimentos e decisões já existentes. Patton *et. al* (1977) reitera em que as avaliações são realizada para corroborar informações sobre as quais não se tem certeza ou, até mesmo, justificar ações. Assim, a percepção dos administradores sobre a validade de uma pesquisa interfere fortemente em sua avaliação. Os administradores que preferem pesquisa ao contrário da intuição são potenciais usuários das avaliações (LEVITON; HUGHES, 1981). Em contrapartida, pesquisas inconclusivas, em que não se tem certeza dos resultados, apontam para o não uso, mesmo que de parte da utilização.

Com relação à apresentação dos resultados das avaliações, o entrevistado do EpR e os avaliadores da FJP destacaram assuntos, momentos e públicos diferentes. Quanto ao assunto,

houve uma primeira apresentação sobre a avaliação de processos, em 2009, e, posteriormente, outra ocasião para a exposição dos resultados da segunda fase em 2010. Quanto ao período da apresentação, observa-se que houve apresentações em governos diferentes: primeiro, na própria gestão do governo em que as avaliações foram contratadas e concluídas, gestão de 2007-2010; e também em uma gestão do novo governo, gestão 2011-2014. Por fim, quanto ao público que recebeu os resultados, nota-se, primeiramente, coordenadores do projeto e pessoas diretamente envolvidas no projeto; em segundo lugar, fóruns secundários, como o Conselho Estadual de Saúde e uma secretaria municipal de saúde, no caso Diamantina, como critério do Banco Mundial; e, terceiro, para os novos gestores da secretaria, que assumiram os cargos no novo governo, como é possível observar nos relatos abaixo.

> No ano de 2009 a gente fez uma avaliação de processo e em dezembro, no final do ano a gente apresentou esses resultados que foram depois discutidos um pouco no início de 2010. Em 2010 a gente fez avaliação de impacto com o campo e no final [dezembro de 2010] a gente apresentou os resultados da avaliação de impacto. Como que foi essa divulgação de resultados? A gente fez uma divulgação inicial para o pessoal do Estado para Resultados. E nessa época também foram convidadas algumas pessoas da SES e também pessoas diretamente envolvidas no programa, pessoas da coordenação do programa, os coordenadores do programa, a equipe aqui de Belo Horizonte que trabalhou com o programa. A gente não fez nenhuma apresentação específica para a Secretaria de Saúde não. (AVALIADOR VV1)

> O final a gente apresentou para a secretaria, a gente, não, a Fundação, apresentou para a Secretaria e o Conselho Estadual de Saúde. Que eu me lembre foi só nesses dois fóruns mesmo, foi para a Secretaria, para o Conselho, e, se não me engano, para uma secretaria municipal, não sei se Diamantina. Porque era critério do Banco Mundial. Quando mudou a gestão [em 2011] a gente tornou a apresentar, mas de uma forma mais breve os resultados da avaliação. A gente apresentou os resultados para os novos subsecretários, como subsídio para os novos gestores. Mas também nunca mais não teve nenhuma outra demanda. (TÉCNICO EPR1)

Entretanto, para alguns integrantes da equipe do projeto Viva Vida, a percepção é de que a apresentação dos resultados tenha acontecido mais *nos bastidores*, para a coordenação do projeto, como relata o entrevistado da equipe do projeto VV1:

> Então, assim, não teve esse espaço de comunicação que virasse para a equipe e falasse "aconteceu isso, o retorno foi esse", não teve uma apresentação para equipe como um todo. Pode ser que tenha acontecido nos bastidores, mas na Fundação com a coordenação, mas eu não me recordo nem se eu estive presente ou se aconteceu. (EQUIPE DO PROJETO VV1, grifo da autora da dissertação)

O processamento de informações, aqui entendido como a tarefa de trabalhar as informações dos relatórios em sínteses e agendar conversas com a Secretaria de Saúde, foi uma função do

EpR. O avaliador VV1 responde como se o cliente e usuário da avaliação fosse o EpR e, não, a equipe do projeto na SES.

> A gente passou tudo para o escritório. A gente sabe que a equipe de lá [equipe do EpR] trabalhou nesses relatórios. Vamos dizer, dos nossos relatórios eles geraram outros relatórios, fizeram reuniões com a secretaria. (AVALIADOR VV1)

> Geralmente você contrata uma pesquisa e vem um calhamaço né, e aí, aproveitando seu título aqui, "vai para a gaveta". O secretário não vai ler um negócio daquele tamanho e mesmo um gestor imediato também poucos eu acho que leriam. Então, a gente tinha essa preocupação de fazer um sumário executivo com as principais conclusões de cada etapa da avaliação. (TÉCNICO EPR1)

A postura dos integrantes do projeto, de acordo com o entrevistado do EpR, foi de que eles já tinham conhecimento sobre as fragilidades do projeto Viva Vida e que já sabiam de todos os resultados apontados nos relatório da avaliação. Tal atitude, como afirma Mokate (2002), reforça a postura defensiva dos usuários da avaliação de ao invés de considerá-la como um instrumento aliado das pessoas para alcançar os objetivos dos programas, classificam-na como algo que irá apontar inconsistências no seu esforço e no seu trabalho diário, levando os resultados da avaliação como críticas individuais à sua maneira de trabalhar. Na perspectiva do gestor do projeto, alguns resultados apontados eram óbvios, como exemplifica, o segundo trecho de entrevista selecionado abaixo.

> A Secretaria de Saúde tem muito uma visão de que já sabiam quais eram os problemas, mas não resolve também. Então pode ter um pouco a ver com isso: "Já sei. Tudo o que você está me falando eu já sei". É um pouco do que a gente escuta agora apresentando resultados de outras pesquisas. (TÉCNICO EPR1, grifo da autora da dissertação)

> E quando a avaliação terminou, por exemplo, e trouxe um resultado que era dos centros terem ficado como predominantemente um equipamento do município-polo e não equipamento da região – que foi um dos resultados que mais chamaram atenção e tal, e a gente concorda absolutamente com eles, aí a gente já tinha inclusive redigido documentos falando: "Mas é obvio. Olha só...". (GERENTE DO PROJETO VV)

> Eu lembro, inclusive, que a coordenadora do Viva Vida, quando a gente apresentou esses resultados, falou: "olha, tudo isso que vocês estão falando aí a gente já sabia, mas fazer avaliação, ter isso por escrito, concretizado nos ajuda muito". (AVALIADOR VV1)

De acordo com Leviton e Hughes (1981), a baixa relevância da avaliação para usuários, sejam eles gestores de programas ou até tomadores de decisões, é fator fundamental ao uso. Cousins e Leithwood (1986) reiteram o grau de importância da avaliação para os públicos que a

receberam, avaliações que refletem o conhecimento do contexto em que elas estavam inseridas e seriam mais bem utilizadas que aquelas afastadas do contexto da organização em que se inserem. No caso das avaliações do Viva Vida, como retratado pelos entrevistados, o momento em que a avaliação foi introduzida não era oportuno aos gestores do projeto. A avaliação, para o entrevistado VV1, até poderia fazer sentido para o EpR, responsável pela contratação, mas não fazia sentido para a equipe do projeto que, no momento de sua realização, enfrentava problemas embrionários do projeto. Para o avaliador VV2, a equipe recebeu o resultado da avaliação como algo inutilizável, já que, na percepção deles, não era o momento de avaliar o projeto. Os resultados ruins, completa, justificam-se pela época errada de se fazer a avaliação na visão dos usuários.

> Eu acho que sempre a avaliação deve acontecer, mas não uma avaliação desse porte naquele momento. Mas é o que eu te falei, tudo depende para quem quer. Quem está contratando podia ser que fazia todo sentido do mundo, para a gente, naquele momento a gente precisava de uma coisa bem anterior mesmo, precisava entender de processo de trabalho, de padronização dos serviços. Outra coisa seria discutir bastante o método mesmo, que na pesquisa isso é muito importante, ter clareza se aquele método é o mais adequado para aquele momento da iniciativa, no caso do Centro. Será que aquilo ali era o que cabia naquele momento? Será que não tinha que estar focando em outra coisa? Assim, adequar mais a avaliação ao momento da política, e não ao contrário, vir com uma proposta que talvez tenha chegado no momento inadequado. A gente não tinha como tomar conta de um que já tinha inaugurado porque tinha que inaugurar outro, ao mesmo tempo não conseguia dar toda força para o que inaugurava porque tinha aquele outro lá para trás cheio de problema. (EQUIPE DO PROJETO VV1, grifo da autora da dissertação)

> O momento da avaliação parece que realmente não foi muito favorável. Na medida em que eles tinham essa convicção de que não era hora, quando se faz uma avaliação que então não mostra impacto, eu acho assim, que eles... Para falar a verdade, eu tenho cá comigo que eles falam "esse resultado é assim porque não era a hora de avaliar". Ponto final. Poe na gaveta e vamos seguir em frente. [...](AVALIADOR VV2, grifo da autora da dissertação)

Em relação à apresentação das avaliações e à receptividade dos usuários quanto às possíveis mudanças, observam-se problemas inerentes, tanto ao formato das apresentações, que priorizava o EpR frente à secretaria de saúde, quanto à credibilidade dos resultados das avaliações frente aos envolvidos com o projeto na SES que, devido à falta de expectativas e baixa importância dado ao processo, consideraram qualquer resultado como inconsistente. Por fim, como resultado dessa análise, a seção seguinte retrata a utilidade da avaliação na percepção de seus vários *stakeholders*.

7.4 Utilização da avaliação do ponto de vista dos *stakeholders*

Para muitos entrevistados, o período imediatamente após a conclusão das avaliações do projeto Viva Vida foi um momento conturbado na secretaria de saúde. Além de mudança física para a Cidade Administrativa, a equipe do projeto vivia um período de alta rotatividade. O próprio projeto sofreu modificações, como a transformação do projeto Viva Vida em programa Viva Vida; a transformação da ação de gestão dos CVVs em processo estratégico; e a criação de um novo projeto estruturador, que também teria o foco na mortalidade infantil, cunhado de Mães de Minas. A utilização da avaliação para a inserção dessas mudanças, no entanto, ainda é ambígua, dado que a visão dos entrevistados é bastante variada.

O projeto Mães de Minas, criado em 2011, trata de um conjunto de ações de saúde voltadas para a proteção e o cuidado da gestante e da criança no primeiro ano de vida, com o objetivo de identificar 100% das gestantes por meio da implantação de um sistema de identificação da gravidez em Minas Gerais (MINAS GERAIS, 2012). Em linhas gerais, o projeto desenvolve ações para a identificação das gestantes mineiras, desde a notícia do início da gestação, e das crianças de até um ano de idade, com a finalidade de identificar possíveis situações de risco e, assim, mobilizar os recursos necessários para o acompanhamento do pré-natal para a gestante e para o cuidado da criança em seu primeiro ano de vida. Para estabelecer a interlocução direta da gestante e família com os serviços de saúde, o projeto conta com a atividade contínua de um *Call Center*, com vistas ao monitoramento do desenvolvimento da gravidez, o nascimento e o desenvolvimento da criança.

É possível verificar diferentes visões sobre a utilização das avaliações (processos e impacto) do projeto Viva Vida Visão sobre a sua utilização, dependendo do tipo de *stakeholder* e do envolvimento deste com a avaliação. Para os avaliadores, é perceptível um desconhecimento quanto à utilização da avaliação como instrumento de mudança no projeto. O avaliador VV2 comenta que não sabe como a discussão foi conduzida na Secretaria de Saúde, dado que perderam o contato com as pessoas do projeto após a conclusão do estudo. O avaliado VV1 até comenta que tem conhecimento de algumas mudanças que ocorreram no projeto após a avaliação, mas não afirma que essas mudanças foram dela derivadas.

> Quer dizer, aí foi até uma questão até que eu discuti com a nova diretora; essa coisa de você fazer uma avaliação de impacto antes da hora tem esse problema. Se a gente

> sabe que a avaliação não deu impacto, mas tendo em vista que o programa estava muito ainda embrionário, no início, a gente não tinha condição de responder isso [se o projeto deveria ser interrompido], até porque o desenho do programa é um desenho muito interessante, quer dizer, tudo indica que ele vai ter resultado. Então, eu sei que por conta disso a gente perdeu o contato e, vamos dizer, eu não sei se de alguma forma isso aí foi discutido na secretaria. Enfim, eu não sei dizer. (AVALIADOR VV2, grifo da autora da dissertação)

> O programa, ele está se renovando sempre né, algumas coisas mudaram, e eu sei que eles criaram algumas coisas em termo de avaliação e monitoramento. Eu acho que isso foi bom também. Acho que eles tinham algumas coisas de solicitações de envio de dados, de informações e eu acho que essa política mudou um pouco lá dentro. Eu sei que algumas coisas novas foram implementadas. Mas eu não tenho contato mais com a coordenação Viva Vida, com as pessoas que estavam lá dentro. (AVALIADOR VV1)

Na visão dos contratantes da avaliação, da coordenação do EpR, o uso da avaliação do projeto Viva Vida foi instrumental, ou seja, as conclusões de fato foram utilizadas para a tomada de decisões ou para a solução de problemas dos programas (LEVITON; HUGHES, 1981). Para o entrevistado da coordenação do EpR, a discussão entre EpR e o alto escalão da Secretaria de Saúde caminharam para a criação de um novo projeto que tentaria sanar alguns pontos de melhoria identificados a partir das avaliações. O desenho do projeto Mães de Minas, de acordo com o entrevistado, seria, assim, fruto da dificuldade de atingir as gestantes de alto risco do estado de Minas Gerais, o resultado da avaliação *chegou bem.*

> No Viva Vida foi legal porque depois a gente teve uma discussão muito boa com o Secretário de Saúde atual, que caminhou até para o Mães de Minas. Então na Saúde também foi importante na hora de desenhar o Mães de Minas. O próprio Secretário reconheceu a necessidade de um foco maior na gestante de alto risco, na busca das gestantes, que na avaliação do Viva Vida não chegavam, a gente não atingia a grávida. Isso foi bom. O debate foi bom, chegou bem. (COORDENAÇÃO EPR2)

Entretanto, na visão de técnicos do EpR, a utilização da avaliação não foi tão certa como afirmado pela coordenação. Em um primeiro momento da entrevista, o técnico do EpR1 entra em contradição com o depoimento da coordenação na medida em que afirma que não ocorreu *nenhuma mudança muito de rumo do projeto em função da avaliação*, embora tenha trazido *resultados relevantes que podem ser trabalhados pela administração*, o entrevistado não tem certeza se os resultados foram usados. Posteriormente, em outra pergunta da entrevista, ao ser relembrado da existência do projeto Mães de Minas, o técnico reafirma a possibilidade de haver interligação entre a avaliação e a criação do projeto, principalmente devido ao problema da focalização identificado na avaliação. Porém, reitera que *pode ter alguma coisa a ver sim, mas não posso afirmar também não.*

> Eu não sei te falar. Eu acho que não teve nenhuma mudança muito de rumo do projeto em função da avaliação. E eu acho que apesar de tudo, ele trouxe resultados relevantes que podem ser trabalhados pela administração. Mas eu acho que tem informações relevantes ali. Eu não sei. Eu estou falando que não foi utilizado, mas de repente mudou isso agora também. Não sei. Como a gente não está acompanhando mais, eu não sei te falar se isso melhorou ou piorou, se eu continuo atendendo quem não devia e atendendo quem devia. Não sei. O Mãe de Minas pode ter alguma coisa a ver com esse problema focalização que foi identificado lá no... Eu acho que tem a ver sim. Pensando bem, acho que... E, lembrando do início lá de 2011, que a discussão do Mãe de Minas passava por isso: "Porque que as mães não estão sendo atendidas?". Então, acho que pode ter alguma coisa a ver sim, mas não posso afirmar também não. (TÉCNICO EPR1)

Na visão do gerente do projeto, o uso foi positivo, de fato, gerou modificações concretas e instrumentais na política de implantação dos centros Viva Vida. Mesmo que nem todas as recomendações das avaliações tenham sido utilizadas, de acordo com o gerente do projeto, o ímpeto da inauguração de CVV foi transformado após a realização das avaliações. Antes, o foco do projeto estaria na inauguração dos CVV, tanto que a ideia original é que todas as microrregiões do estado dispusesse um CCV e, após a avaliação, esse conceito original teria sido modificado. O foco do projeto passou a ser mais no público-alvo do projeto do que na estrutura como explica o entrevistado.

> Olha para você ver como é que isso então gerou uma mudança na política mesmo. De lá para cá a gente freou aquele ímpeto nosso de implantar centro. A gente chegou a falar em ter centro em todas as microrregiões. Daí a gente começou a pensar: "Será que talvez um centro, ele pode atender mais de uma microrregião? Será que realmente precisa de centro ou a gente pode ver um conjunto de serviços, mesmo que fragmentados?". Embora o ideal fosse o serviço mais centralizado e tal, assumir esse serviço, o Estado passar a custear e criar essa sistemática de funcionar com acesso correto. O que vai fazer diferença nesse processo mesmo é eu conseguir fazer com que as pessoas certas cheguem no lugar certo no momento certo. Isso é o conceito da rede basicamente, de atenção, do modelo. Então pouco me adianta ter um centro que é bonito estruturalmente, mas se eu não conseguir garantir esse processo... (GERENTE DO PROJETO VV)

No entanto, ainda na visão do gerente, a reflexão quanto à importância da avaliação do Viva Vida como instrumento de mudança não aconteceu de forma imediata, *teve um efeito meio atrasado*. De acordo com a percepção do gestor, o interesse da avaliação foi surgindo na medida em que ele participava dos diferentes fóruns em que seus resultados foram apresentados, com destaque para a apresentação para o Conselho de Saúde e já no segundo governo por uma iniciativa do EpR. Esse atraso no reconhecimento da utilidade da avaliação, ocorreu devido ao *medo de ser avaliado*, como comenta o gerente.

> Acho que ela valeu. Assim, a duras penas, mas valeu. Ela teve um efeito meio atrasado assim, tipo, talvez se a gente tivesse tido interesse ao longo da avaliação, se

> tivesse um pouquinho menos esse medo de ser avaliado. Eu acho que isso veio a transformar realidade, e aí eu te falo que teve um aspecto muito positivo da avaliação, em dois momentos. Por incrível que pareça, quando da discussão com o Conselho. E num outro momento que o Estado para Resultado já tinha se tornado Escritório, agora já nesse governo. Embora alguns resultados a gente já tinha se dado conta, a gente passou a ter realmente dados que comprovava aquela impressão. Então, isso fez a diferença para a gente. Tanto que eu te falo que de lá para cá a gente já desdobrou essa conversa em algumas situações que retroalimentaram a política. (GERENTE DO PROJETO VV, grifo da autora da dissertação)

No entanto, a visão positiva do gerente do projeto sobre o aproveitamento da avaliação não é compartilhada pelos integrantes da equipe do projeto de forma tão clara. Na percepção do entrevistado VV2, os resultados da avaliação ficaram muito no *nível da coordenação do projeto* e no gabinete da Secretaria de Saúde como um todo, os técnicos mais ligados à implementação e execução do projeto não teriam sequer tomado conhecimento da avaliação. O entrevistado VV1, na mesma posição, também considera que a equipe do Viva Vida nem saberia da existência da avaliação e, devido a isso, acredita que os resultados não tenham sido utilizados.

> Eu acho que ficou muito no nível de coordenação, em nível de gerência. Quando eu falo nível de gerência, é incluindo a coordenação, a gestão de projeto, o gabinete. Eu acho que, do ponto de vista das pessoas que estão ali, as pessoas não tomaram muito conhecimento não. (EQUIPE DO PROJETO VV2)

> Se você perguntar, hoje, para alguém da equipe do Viva Vida ninguém vai nem saber que existiu isso, eu acho. Então, eu acho que a gente pode afirmar com 90% de certeza que nada foi feito em relação aos resultados apurados. (EQUIPE DO PROJETO VV1)

Nota-se, portanto, uma diferença entre a percepção do gerente do projeto e os componentes da equipe do projeto. Enquanto o primeiro salienta o uso instrumental e a relevância da avaliação para modificar os rumos do projeto, integrantes da equipe do projeto sequer tomaram conhecimento dos resultados da avaliação. Um dos motivos do desconhecimento dos membros da equipe sobre a avaliação, de acordo com os próprios integrantes, foi a sua rotatividade no período em que a avaliação foi entregue: *o contexto foi o pior possível*, como relata o entrevistado VV1. A entrega do relatório coincidiu com a mudança para a Cidade Administrativa, troca de coordenação do projeto e uma rotatividade alta por parte da equipe do projeto. O próprio entrevistado VV2 relata que, quando a avaliação foi concluída, ele mesmo não estava mais trabalhando no projeto. Contudo, na visão do entrevistado, era de se esperar que os novos integrantes da equipe recebessem os resultados da avaliação e assumissem.

> Então o contexto foi o pior possível de entrega do relatório. Trocou a coordenação, trocaram os técnicos. Todo mundo que chegava saía. Ficavam seis meses e saía. Foi um período horrível. Então, eu acho que não teve esse retorno de nada. (EQUIPE DO PROJETO VV1)

> Então e não consigo te dizer como é que o retorno disso foi absorvido, se os elementos que foram apontados como problemático no processo de impacto, como é que isso entrou no desenho agora do Viva Vida/Mães de Minas. Não posso te dizer. Infelizmente, o pós-avaliatório para mim foi um período de transição, aí eu já não estava mais na condução do programa. Mas se o Estado investiu recursos para fazer a avaliação é de se esperar, ou, se espera que quem entra, quem assume, receba isso. (EQUIPE DO PROJETO VV2, grifo da autora da dissertação)

Entretanto, ainda que a percepção por parte da equipe dos entrevistados seja que os resultados não tenham sido utilizados, a pesquisa foi útil na visão do VV2. Mesmo com todas as críticas anteriores, o entrevistado VV2 comenta que é importante conhecer o impacto de uma política, além disso, a pesquisa do Viva Vida mostrou que é possível mensurar os efeitos do projeto.

> É engraçado... É o tipo da coisa, o foi útil é difícil eu te dizer, mas, assim, eu acho que do ponto de vista de construção de pesquisa é extremamente útil e necessário. Por quê? A gente precisa saber, a gente está vivendo... Na era moderna da gestão a palavra impacto tem uma força muito grande, então, assim, saber qual é o impacto daquilo que você fez é muito importante saber. Então, variáveis muito importantes em saúde como impacto, como acesso, como efetividade são difíceis de ser mensuradas. Então, assim, é extremamente relevante, e eu acho que a pesquisa mostrou que é possível, de alguma forma, a gente tentar mensurar isso daí. Então eu acho que o mais positivo foi isso, que a gente não esperava que fosse possível fazer, que é mensurar esse impacto. Agora, se é a melhor forma, precisamos fazer novas pesquisas para ver se a gente consegue aproximar mais. (EQUIPE DO PROJETO VV2)

Na mesma linha de argumentação do integrante da equipe do projeto, para o avaliador VV1, a realização da avaliação do Viva Vida, por si só, trouxe benefícios para os diferentes envolvidos na avaliação: primeiro para o desenvolvimento da Fundação João Pinheiro quanto a realizações de programas governamentais, dado que tem como atividade essencial a pesquisa de políticas públicas; do ponto de vista do EpR (chamado pelo entrevistado de Escritório) e da SES, a geração de um conjunto de informações quantitativas e qualitativas concretas sobre o projeto que podem ser utilizadas, tanto para repensar as diretrizes da política do Viva Vida, quanto para de fonte de consulta e registro e, por fim, para o Banco Mundial, pela capacidade de reconhecer o projeto como uma iniciativa séria, como retratado no primeiro trecho de entrevista abaixo. Além disso, ressalta que a própria dificuldade em avaliar, de encontrar os métodos mais apropriados, de localizar beneficiários para a formação de um banco de dados, acarreta pontos de reflexão que são positivos para o desenvolvimento

do setor público como um todo. O entrevistado do EpR completa que, ainda que a informação gerada não tenha acometido ao uso, o processo foi importante para gerar uma *curva de aprendizagem* no governo quanto ao assunto avaliações, atualmente as pessoas que participaram têm mais condições de escolher o tipo correto e discutir melhor, como argumenta no segundo trecho selecionado.

> Eu destacaria essa importância de fazer avaliação, eu acho importante para o Escritório, importante para a Secretaria e importante para a Fundação. Para a Fundação, porque a Fundação vive disso, vive de pesquisa, vive desse tipo de trabalho e que foi interessante para o crescimento da equipe. Para a equipe da Fundação eu acho que significou um crescimento muito grande, porque a gente cresceu muito em termos de realizar avaliações desse tipo, dessa natureza. Eu acho que os pesquisadores aqui cresceram muito com essa avaliação. Foi interessante para o Escritório porque eu acho que eles puderam ter informações para reavaliar imagino que algumas diretrizes da política. Para a Secretaria eu acho que foi interessante para ter as informações, principalmente em relação à avaliação de processo, do que poderia ser melhorado do programa. Eu acho que foi uma experiência muito interessante. Para o Banco Mundial foi bom também porque, afinal de contas, o Banco pode entender e ver que o programa foi instalado, que ele é um programa sério. Mas ela foi interessante por alguns aspectos, foi interessante para ver um pouquinho dessas dificuldades de ter o que avaliar em termos de dados, de ter informações quantitativas que nos permitisse uma avaliação. Então é bom pensar isso, é bom saber que não existem dados, é bom saber que não tem informações para também pensar o que fazer para suprir essa necessidade. (AVALIADOR VV1, grifo da autora da dissertação)

> Eu acho que tem uma curva de amadurecimento nesse processo todo. Então, assim, foram os primeiros que a gente fez e eu acho que acumulou conhecimento com isso. Hoje a gente tem muito mais condição de discutir que tipo de avaliação serve para cada projeto, quais são os pré-requisitos para uma boa avaliação. (TÉCNICO EPR1, grifo da autora da dissertação)

Contudo, ainda que os relatos do avaliador e do técnico do EpR apontem para o aprendizado adquirido no processo, relacionado ao uso processual da avaliação que, como afirmado por Patton (2008), acontece somente quando mudanças da forma de pesar resultantes da participação no processo de avaliação acontecem nos usuários, a percepção de que a equipe gestora do projeto Viva Vida deveria ser mais inserida nas decisões e no dia a dia da avaliação foi unanime entre os entrevistados. Quando questionados pela pergunta *o que fariam diferente caso a avaliação do Viva Vida fosse realizada hoje?* –, os entrevistados direcionaram sua respostas para a inclusão dos usuários, tanto do ponto de vista dos avaliadores, de incentivar a participação e incluir os usuários na decisão de avaliar, quanto da própria equipe, de procurar participar mais e se fazer incluída. Assim, valorizar mais a equipe da SES e tentar compartilhar mais decisões da avaliação, de forma que a mesma seja realizada com o apoio total da área avaliada, com um consenso geral entre as partes, como uma demanda construída

em conjunto, foi um ponto comum nas entrevistas de avaliadores, gestor e equipe, como demonstram respectivamente os relatos.

> O que eu faria de diferente é exatamente assegurar que a política que fosse avaliada, ela tivesse o apoio total, que o pessoal, por exemplo, da secretaria implementadora, que achasse que era a hora, o momento de avaliar ou pelo menos, vamos dizer assim, ao menos que os argumentos dele não fossem convincentes. "Tá na hora de avaliar?" "Qual o tipo de avaliação mais apropriada?". A gente não pode chegar assim com uma demanda pronta: "vamos fazer uma avaliação de impacto". Não! "Vamos fazer uma avaliação de que programa e que tipo de avaliação que é melhor para aquele momento do programa?" (AVALIADOR VV2, grifo da autora da dissertação)

> Eu não vejo como construir um processo mais maduro, se não tiver mesmo na cabeça do gestor, na cabeça dos técnicos, há clareza de que a gente não pode continuar definindo o praxismo. É o ponto de partida. Se "teve" algumas coisas que não foram do jeito ideal nesse processo, elas tiveram muita ligação com a nossa visão sobre a avaliação. (GERENTE DO PROJETO VV, grifo da autora da dissertação)

> Eu acho que primeira coisa eu tentaria envolver a equipe e mostra que a equipe é coresponsável dessa avaliação. Porque senão vai acontecer isso, vai ser um papel a mais que vai está lá. Então, a partir do momento que eles entendem que aquilo ali eles estão fazendo com a finalidade para mudar o trabalho deles, para melhorar, eu acho que tem a sensibilização para pesquisa dar em alguma coisa, para a avaliação acontecer, senão vira uma coisa meramente *pro forma*, que é interesse pontual de fulano ou siclano, não é uma coisa da coordenação, da instituição. Eu acho que seria isso. (EQUIPE DO PROJETO VV1, grifo da autora da dissertação)

Como considerações finais da avaliação do projeto Viva Vida, atenta-se para o seu contexto inicial conturbado, marcado pela resistência dos envolvidos na SES quanto ao objeto e ao momento em que o estudo avaliatório foi realizado. Em função desse contexto, fica evidente a indiferença da SES com os resultados que poderiam ser descobertos a partir da avaliação, assumindo a avaliação como um processo imposto, limitando-se ao fornecimento de dados regulado pelo EpR. Quanto à execução do estudo avaliatório, notam-se diferenças metodológicas, como a realização de uma avaliação de processos, adicional à avaliação de impacto, e dificuldades operacionais ao longo da coleta de dados o que, na visão de alguns atores, demonstrava que o projeto não estava pronto para ser avaliado. Por fim, em relação à utilização, nota-se um uso circunscrito ao alto escalão do EpR e da SES, já que implementação do Mães de Minas, derivada da avaliação do Viva Vida, só pôde ser percebida pelo gerente do projeto e pelo coordenador do EpR. Os demais entrevistados emitiram a percepção de que ou desconheciam a aplicação dos resultados dos estudos ou não arriscariam a emitir opiniões mais concretas, devido ao afastamento do mesmo, como fizeram os avaliadores, os integrantes da equipe do projeto e o técnico do EpR envolvido.

8 AVALIAÇÃO DO PROJETO GESTÃO INTEGRADA DE AÇÕES E INFORMAÇÃO DE DEFESA SOCIAL

A avaliação do projeto estruturador Gestão Integrada de ações e informação de Defesa Social constitui uma das avaliações do pacote organizado pela FJP e EpR em conjunto com o Banco Mundial e, portanto, contou com a participação de consultor contratado e avaliadores da FJP. Para a pesquisa empírica deste trabalho, contou-se com a participação de apenas quatro entrevistados diretos da avaliação do projeto, o que compreende: gerente do projeto, avaliador da FJP, consultor externo contratado pela FJP e o técnico do EpR que acompanhou o projeto de avaliação. Também buscaram-se percepções de entrevistados da coordenação do EpR e da coordenação da FJP. De início, dada a ausência de potenciais entrevistados para este trabalho, é possível verificar a participação reduzida de diferentes *stakeholders*. Destaca-se também que essa avaliação foi a única, dentre as pesquisadas neste trabalho, em que o consultor externo se mostrou solícito à realização de entrevista.

O capítulo está organizado em quatro seções que remontam ao histórico da avaliação do projeto, a partir da percepção dos entrevistados, desde a concepção da avaliação e expectativa inicial, até a sua conclusão, com alusão ao tipo de utilização ocorrido. Na primeira seção apresenta-se o contexto em que a avaliação foi inserida, seguida pela implementação e envolvimento dos atores na execução da avaliação. A terceira e quarta seções tratam, respectivamente, dos resultados e da utilização da avaliação.

8.1 Contexto geral da avaliação

Os estados brasileiros, assim como em outros países do mundo, contam com sistemas de segurança pública compostos por forças polícias independentes, nos quais a polícia militar responsabiliza-se pelo policiamento ostensivo e a polícia civil cumpre o papel de polícia investigativa. No entanto, a falta de compartilhamento de informações entre as organizações policiais, a incompatibilidade de áreas territoriais de atuação e a inexistência de planejamento conjunto de ações causam obstáculos importantes à efetividade da ação policial (SAPORI; ANDRADE, 2008). O projeto estruturador Gestão Integrada de ações e informação de Defesa Social, objeto desta análise, é o componente responsável pela inversão dessa lógica perversa causada pela independência entre a Polícia Militar e a Polícia Civil e pretende, por meio de diferenciadas ações, implementar a política de integração de defesa social observados os

princípios constitucionais e respectivas competências de cada ente. Em linhas gerais, o projeto visa responder a duas situações evidenciadas em relação à segurança pública em Minas Gerais: aumento dos índices de criminalidade e baixa efetividade do aparato estatal em manter a ordem pública (SOARES; SOUZA, 2010).

O projeto estruturador, quando implantado inicialmente em 2003, estava inserido em uma política governamental de segurança mais abrangente, como mostram Sapori e Andrade (2008), fundamentado com a criação da Secretaria de Estado de Defesa Social (SEDS) em 2003 e sob sua responsabilidade foi colocada a coordenação operacional das atividades das Polícias Civil (PCMG) e Militar (PMMG), do Corpo de Bombeiros Militar (CBMG), da Defensoria Pública e da Subsecretaria de Administração Penitenciária. A partir disso, o modelo de integração concebido para Minas Gerais pautou-se na constituição dos centros integrados de comunicações como organismos estratégicos para aproximação das polícias civil e militar, tendo como estratégias principais a integração das informações, das áreas geográficas de atuação e seu planejamento operacional (SAPORI; ANDRADE, 2008). Pensando nessas estratégias, foram propostos três projetos: na área informacional – Sistema Integrado de Defesa Social (SIDS); no que tange à integração geográfica – Áreas Integradas de Segurança Pública (AISP); e para o planejamento operacional – Integração da Gestão em Segurança Pública (IGESP). De acordo com o gerente do projeto, em conjunto, essas três grandes estratégias, que, na verdade, constituíam outros projetos independentes, provocavam vários questionamentos pelos diferentes *stakeholders* que participavam do projeto. Não existia, na visão do gerente, um consenso quanto ao funcionamento e efetividade.

> Havia um problema de desalinhamento de visões estratégicas. Ou seja, a PM pensava uma coisa a respeito do projeto, a civil outra, a secretaria outra. Assim, não havia um consenso muito claro do que se pretendia com esse projeto, e mais, com a política. Muito provavelmente porque foi uma política, a política de integração, ela foi uma política construída e pensada a partir da cabeça de poucas pessoas que estavam em 2002, que eram as cabeças pensantes da parte da civil, outras da PM, outras da academia. (GERENTE DO PROJETO IGESP)

É possível observar que o gerente do projeto percebe o projeto estruturador como uma política geral de integração, desenhada por poucas pessoas do governo e da academia, composta por diferentes projetos, quais sejam SIDS, AISP e IGESP, projetos esses que apresentavam ritmos de desenvolvimento e abrangência diferenciados. Este trabalho seguirá o entendimento do gerente da política de integração, também responsável pelos três projetos subordinados à política, uma vez que, quando se decidiu avaliar o impacto da política não foi possível avaliar

todo o projeto, mas parte dele, como comentado à frente pelos entrevistados. A ideia inicial da avaliação seria, assim, realizar uma avaliação de impacto de algum projeto da área de defesa social, sem saber ao certo qual deles estaria mais preparado para passar pela análise. O relato do consultor externo, contratado pela FJP, para realizar a avaliação de defesa social conta como aconteceu o convite para participar da iniciativa e como o desenho da avaliação convergiu para a avaliação do IGESP em apenas uma reunião.

> Que eu me lembro, a avaliação de políticas públicas, ela queria avaliar grandes projetos que o Estado tinha lançado em áreas-chave, e uma delas era segurança pública. Já havia uma literatura anterior mostrando evidências de que integração, pelo menos, ações coordenadas da policia civil e da policia militar reduziam criminalidade em vários pontos do mundo. Tem experiência em Nova York, em Bogotá se teve uma experiência similar. Então, esse era um projeto-chave do Governo, vamos dizer assim. Segurança pública é sempre um assunto muito importante. (AVALIADOR IGESP)

> O contato inicial foi a Fundação João Pinheiro. Mas a reunião inicial foi com o pessoal do Estado para Resultados. Foi uma reunião inicial sobre segurança pública, sobre medidas que estavam sendo adotadas e que eram uma das prioridades, e quais seriam as alternativas de avaliação. Entre idas e vindas dessa conversa inicial tinham programas, algumas ideias em termos de disponibilidade de dados, facilidade de avaliação num tempo hábil, e dada também, de certa forma, a prioridade política do governo, qual que despertava maior interesse dele, acabou-se convergindo para a avaliação do IGESP. (CONSULTOR IGESP)

O IGESP é um modelo gerencial que busca implementar uma interlocução recorrente entre os órgãos de defesa social e viabilizar as ações integradas entre as polícias no combate direto à criminalidade. Assim, visa promover o aproveitamento das informações de inteligência atualizadas e precisas para os agentes de segurança (SAPORI; ANDRADE, 2008). Adaptado de modelos de gestão do trabalho policial de Nova York e Bogotá, o projeto se desenvolve por meio de reuniões de trabalho em cada uma das AISP, denominadas reuniões de apresentação do IGESP. Nessas reuniões, os gestores e pessoal operacional das polícias civil e militar, em conjunto, trocam experiências, compartilhando informações e discutindo possíveis soluções para os problemas de segurança pública daquela região. A partir dos encontros e de posse das informações, é de se esperar que as forças policiais tenham maior propensão a implementar estratégias de combate à criminalidade mais eficazes, com a alocação rápida, sincronizada e focalizada de recursos. O projeto pode ser entendido, assim, como um modelo de organização e gestão do trabalho policial para integrar ações de segurança advindas de um

modelo de policiamento orientado por informações[8]. Nas palavras do gerente do projeto, o funcionamento do IGESP ocorre do modo como se vê nesta entrevista.

> Na defesa você divide territórios em regiões integradas. O que seria isso? Metodologia e IGESP? Era juntar as pessoas da área de segurança pública, que suas atividades estavam envolvidas com esse problema, com esse fenômeno, levantar o maior número de informações possíveis de estatísticas e tal, as pessoas vão sentar, vão analisar o contexto da violência ali, dos problemas que eles estão tendo e vão traçar planos de ação. (TÉCNICO EPR2)

Sapori e Andrade (2010) apresentam, no entanto, as dificuldades de legitimação do projeto frente aos próprios participantes. Para os autores, embora a ideia do projeto de colaboração e reciprocidade entre o policiamento ostensivo e o investigativo pareça ser teoricamente consistente, não seria possível traçar, de imediato, uma mudança de cultura na interação das forças policiais. Para alcançar o seu objetivo, o projeto teria que lidar com os constrangimentos institucionais existentes entre as duas esferas da polícia, o que inclui a superação de preconceitos e estereótipos. Nesse sentido, difundir uma crença de que a unificação das polícias ou a integração de suas ações levaria a uma maior eficiência da ação policial não foi algo trivial. Ainda assim, o IGESP foi sendo progressivamente expandido, sendo inicialmente implantado, em 2005, como um projeto-piloto na capital do estado, até incluir 56 municípios no ano de 2008 (FUNDAÇÃO JOÃO PINHEIRO, 2010c).

O desafio do IGESP de integrar duas organizações, com arranjos institucionais consideravelmente distintos, somado à implantação contínua do projeto instituiu, assim, dúvidas sobre a efetividade do projeto, como comentado pelo gerente. Para o gestor, em 2007, a política de integração precisava ser analisada e questionada para saber se o projeto estava no caminho certo.

> O IGESP começou em 2005 a implantação a partir da capital, começou com um piloto e depois houve uma expansão já em 2006 para seis municípios do interior. Então, o IGESP estava inserido dentro de um contexto maior de uma política de integração que já tinha caminhado um tanto, mas que precisava ser analisada no sentido assim, "até que ponto a direção é essa?". Meio que ela estava sob análise mesmo, assim: porque que algumas coisas não avançaram? O que que está acontecendo? Então 2007, foi um ano que foi provocado um momento de parar e analisar, fazer um balanço do que que tinha avançado, do que não tinha avançado e para onde ainda você tinha que ir. (GERENTE DO PROJETO IGESP)

---

[8] Para maiores informações sobre o projeto, ver Sapori e Andrade (2010).

O avaliador do IGESP e o consultor externo concordam com a afirmação do gerente do projeto, na medida em que colocam em xeque a legitimidade do projeto frente aos envolvidos. De acordo com os relatos abaixo, havia uma atitude de aversão à lógica de integração da política, dado que as ações do projeto exigiam uma quebra do paradigma da independência das polícias e muitos receavam a perda de espaço de uma instituição frente à outra.

> O próprio Estado para Resultados, que estava intermediando a coisa com a secretaria, nos passou a opinião de que "É muito bom que vocês estejam presenciando isso porque existe um impasse dentro da polícia de trabalhar junto" por causa de cultura institucional, do nível de interação das policias civil e militar que não era usual, e até uma reestruturação meio hierárquica entre as duas forças, e uma mudança de atitude em relação ao que era tradicional dentro da polícia. Então acho que existiu uma resistência entre determinados grupos no sentido de "Isso não vai mais funcionar. Isso é uma bobagem. Isso aí não está tendo nenhum feito". E teria sido muito bom a gente poder sugerir para o próprio comando da polícia, para que tivesse uma visão mãos objetiva que estava surgindo a partir daí. (CONSULTOR IGESP, grifo da autora da dissertação)

> Eu acho que, dentro da estrutura de projetos que o Governo Aécio Neves lançou, esse era um dos mais óbvios a ser avaliado até porque muito se criticava a respeito dessa integração. Muitas pessoas falavam: "Isso não vai dar em nada. Isso não funciona", inclusive, até mesmo, dentro das próprias instituições, dentro das mesmas instituições envolvidas, a polícia civil e a polícia militar. É claro que a gente sabe que historicamente, como foram duas instituições que cresceram separadas, sempre há ali um medo, um receio, digamos assim, da perda de espaço de uma para outra. Não é fácil você integrar instituições. E, de fato, as instituições, elas não estão integradas, eu acho que talvez esse seja um termo forçoso. Integradas elas não estão porque elas não se tornaram uma instituição, elas passaram a fazer reuniões para coordenar ações. Então, o grande mérito é colocar os indivíduos das duas instituições lado a lado, um falar para o outro os problemas que eles estão tendo naquela localidade, e o que a outra instituição pode contribuir comigo e, vice versa, para jutas, elas melhorarem os indicadores de criminalidade. (AVALIADOR IGESP)

Observa-se, a partir dos relatos, que uma possível função da avaliação seria justificar, de forma objetiva, a importância do projeto para os atores que não confiavam na sua proposta e não acreditavam que a integração poderia gerar resultado. As entrevistas sugerem, portanto, o emprego da avaliação do IGESP como uma saída ao descrédito do projeto frente aos seus participantes, servindo como instrumento de persuasão política. Contudo, mesmo que os pesquisadores ressaltem a importância da avaliação de impacto do IGESP para legitimar o processo e ganhar adeptos ao projeto, o gerente do projeto não entende que esta seja a justificativa para a realização do estudo de efetividade do projeto. Ainda que não tenha certeza, o gestor supõe que o projeto tenha sido selecionado para a avaliação porque constituía, na época, *a menina dos olhos* da Secretaria de Defesa Social devido a dois fatores: primeiro, por causa da sua natureza de organização processual; e segundo, porque o projeto

tinha um orçamento relativamente baixo frente às demais políticas de integração, o que possibilitava a sua expansão para todo o estado.

> Por isso que eu falo, porque escolher de tudo o IGESP na Avaliação de impacto? Aí é uma suposição: o IGESP era a bola da vez, ele virou a menina dos olhos. Assim, intuitivamente no alinhamento: "Esse projeto é um projeto aparentemente de consenso, todos os atores querem que ele chegue a todos os lugares porque isso não é caro, isso é método, é método que organiza os processos". Até o Secretário, eu lembro do Secretário falando assim: "O IGESP é o único projeto que a gente pode fazer a política chegar no Estado inteiro da política de integração, porque o resto tudo demanda muito dinheiro". (GERENTE DO PROJETO IGESP)

Entretanto, mesmo tendo o IGESP como uma prioridade, o gerente do projeto alega que a decisão de avaliar o projeto não foi uma deliberação colegiada. Mais uma vez supõe que a escolha tenha sido mais uma decorrência das discussões entre os secretários de defesa social e do EpR. Não diferente dos demais projetos avaliados, o gestor só foi avisado que a avaliação iria acontecer, *o que não significa que a SEDS tenha sido comunicada* argumenta o gestor, como exemplifica o trecho abaixo.

> Então, essa coisa também dele ser eleito porque a pesquisa de impacto, porque que ele é prioritário não foi assim: "Gente, quem que é a nossa prioridade? Ah, é o IGESP". Acho que foi muito a condução do Secretário de Defesa com o Secretário do Estado para Resultados. Estou supondo. Porque essa coisa de quando se decidiu que vai fazer a Avaliação de Impacto do IGESP, eu fui comunicada, mas isso não significa que a SEDS tenha sido comunicada. (GERENTE DO PROJETO IGESP, grifo da autora da dissertação)

Na percepção do gerente, a pesquisa foi muito bem recebida, pois, independentemente da maneira que foi negociada, a avaliação faria o papel de uma *bússola de orientação* para o um projeto como o IGESP. Tal analogia converge para o conceito que avaliar significa atribuir um valor, determinar a valia de algo e, assim, necessariamente, trazer uma medida de aprovação ou desaprovação das questões avaliadas (ARRETCHE, 1998). Como adequadamente completam Aguilar e Ander-Egg (1995, p. 31), as avaliações são destinadas à "[...] identificar, obter e proporcionar de maneira válida e confiável dados e informações suficiente e relevante para apoiar um juízo sobre o mérito e o valor dos diferentes componentes de um programa". A avaliação funcionaria, dessa maneira, como um guia que, nas palavras do gestor, seria como *uma bússola de orientação*, assim como qualquer pesquisa que aparecesse naquele momento do projeto seria bem recebida.

> Porque como era uma equipe muito nova, que lidava com o novo, no fundo no fundo a gente via aquilo como uma bússola de orientação. Todo mundo ansiava. Se você

> falasse assim: "Eu quero uma pesquisa semestral, anual". "Eu quero!". Porque, assim, a gente estava lidando com uma realidade assim, construindo, você está construindo uma coisa que você não tem certeza do que está funcionando, o que não está, dos problemas. Então todo mundo ficava era assim: "Nó, pesquisa...". Isso do ponto de vista de equipe. (GERENTE DO PROJETO IGESP, grifo da autora da dissertação)

Por parte dos demais envolvidos no projeto, a receptividade quanto à avaliação não foi tão positiva, na visão do consultor e do técnico do EpR que acompanhavam a avaliação. Como já existia certa incredulidade sobre a proposta geral do projeto por parte de alguns grupos, a avaliação também não seria algo tão bem recebido em um contexto que envolvia uma cultura institucional forte. Para o consultor da avaliação, o sentimento de aversão ao estudo avaliatório seria, até certo ponto, natural, frente às descobertas que a mesma poderia trazer. E completa que, mesmo causando desconforto, as avaliações devem ser realizadas. Já na visão do entrevistado do EpR, as avaliações causavam resistência por terem sido encomendadas por um órgão externo à defesa social, assumindo, assim, um caráter fiscalizatório.

> Talvez se tivesse alguma resistência, talvez em alguns níveis alguma atitude de descrença de certa forma, o que é natural. Acho que um projeto que mexe em estruturas predefinidas, que envolve ali relação de poder, é natural que qualquer avaliação com o resultado seja positivo ou negativo vai despertar um desconforto. Mas não é por isso que a avaliação não deve ser feita, né? Para apresentar dados objetivos que ao menos instruam um pouco mais um debate material, diferente daquele que emite a opinião de pessoas com diferentes interesses e não leva a nada. Acho que também seria em geral, mas acho que talvez em algumas dimensões houvesse essa descrença, mas acho que é natural. (CONSULTOR IGESP, grifo da autora da dissertação)

> Porque o órgão também, ele sentia assim com um pouco de medo. "Você está vendo o Estado dar resultado aqui, está trazendo consultor externo, está querendo avaliar, está querendo avaliar, está querendo informação". Então está pensando assim; "Nossa, eles estão querendo me controlar, aditar, fiscalizar!". Eu acho que eles tinham um pouco de resistência. (TÉCNICO EPR2, grifo da autora da dissertação)

Quanto a esse tema, Mokate (2002) ilustra a visão de *monstro* da avaliação pelos gestores justamente por enxergar a avaliação como uma pesquisa muito sofisticada, fiscalizatória ou, até mesmo, imposta. Contudo, como complementa a autora, tal percepção é devida à falta de interação dos gestores dos programas, que não se enxergam como aliados dos estudos avaliatórios (MOKATE, 2002), como será discutido na seção seguinte. Nessa oportunidade, remonta-se o envolvimento dos atores na implementação da avaliação do IGESP, bem como o seu planejamento e sua implementação.

8.2 Implementação do processo de avaliação e envolvimento dos atores

Shulha e Cousisn (1997) afirmam que desenvolver avaliações úteis vai além da simples combinação de métodos de avaliação, exige uma preocupação com fatores contextuais por parte dos avaliadores. Cabe recordar que diferentes autores ressaltam o papel influenciador do avaliador na promoção e cultivo do uso, como asseguram Patton (2008) e Johnson *et al.* (2009). Mesmo, que haja um consenso quanto à responsabilidade dos avaliadores de extrapolar o simples auxílio aos usuários sobre a abordagem da avaliação, muitas avaliações são realizadas de forma insulada e isoladas de seus usuários. No caso do IGESP, fazendo uma análise das entrevistas, logo se percebe que a interação entre o avaliador e consultor com as pessoas da Secretaria de Defesa Social foi mínima, sendo todo o processo mediado pelo EpR.

Como todo o relacionamento aconteceu com o EpR, tanto consultor como avaliador, reiteram, em diferentes momentos da entrevista, o papel essencial do EpR na condução do processo, como exemplificado no relato abaixo.

> O Estado para Resultados foi importantíssimo, principalmente na interlocução com a SEDS, com a polícia militar e polícia civil. Eu acredito que a nossa vida tenha sido muito facilitada pela presença do Estado para Resultados, que é quem fazia a ponte. De fato, a gente conseguiu um acesso mais fácil a essas instituições porque a gente tinha o apoio e a presença das pessoas do Estado para Resultados envolvidas na busca por informações também. Então, de fato, acho que facilitou a nossa entrada na SEDS, facilitou a nossa entrada nas polícias civil e militar sem dúvida. (AVALIADOR IGESP)

Ademais, ainda em relação ao EpR, com respeito à atitude das pessoas que lá trabalhavam e que interagiram com o consultor contratado, o consultor ressalta o profissionalismo dos representantes do EpR no trato de assuntos que tem a ver com avaliação. A postura das pessoas, comenta, foi bastante técnica e objetiva quanto às avaliações, não foi observado nenhum indício de intervenção política ou direcionamento nos resultados da pesquisa. De acordo com Alkin e Coyle (1988), não é incomum que os encomendantes de uma avaliação, como se encontram envolvidos nas questões políticas das burocracias, estejam mais direcionados a manipular os resultados de uma avaliação de forma intencional como, por exemplo, desviar informações dos relatórios de forma intencional para divulgação, atitude essa que é entendida pelos autores como abuso da avaliação (*evaluation abuse*). E, de fato, como argumentam Bamberger, Rugh e Mabry (2011), as intenções políticas podem prejudicar o trabalho dos avaliados quando são pressionados a ter uma conduta inadequada para um

avaliador, seja por necessidade de manipulação de dados ou por imposição de que os resultados sejam positivos. No entanto, no caso da avaliação do IGESP, o consultor relata que ficou tão bem impressionado com a maneira que as avaliações foram tratadas pelos representantes do EpR que chegou a comentar com colegas de profissão, evidenciando, portanto, a percepção da postura do EpR como uma exceção à regra, como exemplificam os trechos selecionados.

> Até fiquei bastante impressionado com a atitude do Estado, com quem eu interagi mais, que foi o Estado para Resultados em relação à avaliação. Então, era realmente uma atitude de "Nós que temos que saber o que estava acontecendo com o projeto". Não é uma atitude de "Queremos uma bandeira para mostrar que o programa está funcionando". Então, não houve em momento algum, digamos, nem uma ponta de sugestão de que gostariam que os resultados fossem positivos ou negativos, ou, "Gostaríamos disso", ou, "Não podemos dar esses dados". Então essa postura é muito condizente com a proposta de uma avaliação, e isso eu fiquei muito impressionado na realidade. A coisa que mais me chamou atenção foi isso, o profissionalismo das pessoas com quem eu interagi, profissionalismo e na verdade o nível de qualificação das pessoas com quem eu interagi no Estado para Resultados e ausência de qualquer sombra de pressão política sobre resultado. É a ideia de "Queremos ver o que está acontecendo". Eu fiquei muito impressionado, fiquei impressionado a ponto de comentar isso com colegas aqui que trabalham com avaliação em vários lugares. Então eu achei isso realmente muito louvável e raro no Brasil, infelizmente. (CONSULTOR IGESP, grifo da autora da dissertação)

Os elogios ao EpR, no entanto, não são estendidos à SEDS, dada à deficiente participação dos técnicos do projeto para a gestão integrada na avaliação. A falta de proximidade e de contato com a avaliação pode ser demonstrada pela dificuldade de encontrar possíveis entrevistados na SEDS para a realização desta pesquisa empírica. Dentre vários envolvidos no projeto por parte da SEDS, apenas o gerente do projeto participou da entrevista e confessou que não poderia indicar mais pessoas, pois elas não se envolveram com o estudo avaliatório do IGESP. Na versão do entrevistado do EpR, existia uma *distância muito grande* entre o pesquisador e a política avaliada. A interação com a Secretaria de Defesa Social, nesse sentido, foi restrita a ponto de os avaliadores não conhecerem completamente o funcionamento e o contexto, como comenta o entrevistado no trecho abaixo.

> O Estado para Resultados na época que fazia essa mediação e tal, mas eu senti, pelo menos nas avaliações que eu acompanhei, que tinha uma distância muito grande. Tinha que fazer a metodologia, pensar política, mas os pesquisadores que estavam desenhando a avaliação, eles conheciam pouco, eles tiveram pouco conhecimento. Eu acho que devia ter tido mais interação mesmo com os gestores do projeto para conhecer a realidade. (TÉCNICO EPR2, grifo da autora da dissertação)

A comunicação com a instituição avaliada que, na visão de Cousins e Leithwood (1986), deve ser permanente durante toda a pesquisa, deu-se em dois momentos distintos: no início da avaliação, com o comunicado de que ela iria ocorrer e com poucas conversas sobre a dinâmica do projeto; e, posteriormente, na fase de apresentação dos resultados, como comentado pelo próprio consultor da avaliação. No seu entendimento, enquanto o contato com a SEDS foi de cunho operacional, direcionado à necessidade de entender o programa, a coordenação do EpR era, de fato, a quem se reportavam já que, desde o início a demanda partiu do EpR e, não, da secretaria de defesa. Nesse sentido, o consultor evidencia que a relação com a SEDS *foi um pouco mais restrita*. Na percepção da coordenação da FJP, uma das fragilidades do grupo de avaliações foi justamente a falta de interação contínua entre avaliador e os envolvidos no projeto avaliado. A incoerência estava justamente no contato reduzido das avaliações, primeiro no início e, depois, já na conclusão do processo, já com o relatório em mãos, como se retrata no segundo trecho de entrevista selecionado.

> A nossa interação com a SEDES foi um pouco mais restrita. "A gente quer esse dado, a gente precisa desse tipo de dado. Tem esse dado aí?". Então, quer dizer, desde o começo a demanda da avaliação partiu do Estado para Resultados. Quer dizer, a demanda por avaliação não partiu da secretaria no primeiro momento. A gente teve algumas reuniões na SEDES até para esclarecer alguns aspectos mais operacionais do programa que a gente não entendia muito bem e foi lá para conversar com o técnico para explicar a dinâmica de determinadas coisas, exatamente como é que acontecia e tal. Depois, no final das contas, no processo de divulgação dos resultados, no final das contas o ápice, digamos, do processo, foi quando a gente apresentou para o secretário de defesa social. (CONSULTOR IGESP, grifo da autora da dissertação)

> E eu acho que outro problema é que os projetos foram conduzidos completamente a distância, sem nenhum a interação com os técnicos trabalhando nos programas que estavam sendo avaliados. Você foi lá uma vez, colheu os dados, foi embora e voltou com o relatório. (COORDENAÇÃO FJP, grifo da autora da dissertação)

A dinâmica da troca de dados e de informações, que comporiam a avaliação do IGESP, exemplifica a infrequência da comunicação entre avaliadores e SEDS. Ainda, de acordo com o consultor da avaliação, todos os dados, apesar de serem originários da SEDS, chegavam via EpR. Ou seja, os dados eram produzidos pela defesa social, em seguida, eram repassados ao EpR e só depois eram encaminhados aos avaliadores da FJP, como relatado no trecho abaixo.

> O pessoal da Fundação João Pinheiro e o próprio avaliador da FJP até naquele momento que entraram mais em contato, mas foi tudo via Estado para Resultados. Uma grande parte dos dados, de criminalidade, número de municípios, número de viaturas, diferentes programas que estavam sendo implementados com esse nível, número de prisões e coisas desse tipo, isso tudo, veio da Secretaria de Defesa Social para o EpR. Acho que obviamente existia uma percepção de que a avaliação é uma

> prioridade do Estado, isso acho meio personificado no Estado para Resultados, e realmente havia uma abertura muito grande nesse sentido. (CONSULTOR IGESP)

Ainda que a associação da avaliação do IGESP ao EpR e, não, à SEDS pareça incoerente, segundo a percepção de alguns entrevistados, como o técnico do EpR e o próprio gerente do projeto, a secretaria de defesa tinha outras prioridades que não se relacionavam à avaliação. Um primeiro elemento para se pensar diz respeito ao enfoque em atividades, característica essa que tradicionalmente acomete a gerência de programas sociais, como afirma Mokate (2002). Igualmente como nos demais casos aqui estudados, observa-se uma equipe de projetos orientada ao cumprimento de atividades e processos e o produto máximo do trabalho são entregas de serviços e, não, de resultados (MOKATE, 2002). Adjunto a esse estilo gerencial, que somente requer informações de monitoramento, as equipes não tenderão a contar com incentivos e, muito menos, interesse para realizar e encarregar-se de avaliações de impacto. Nesse sentido, a gerência e a equipe dos projetos se envolvem com uma administração focada em atividades facilmente controladas por eles e se afastam do tema e da importância da avaliação. Como exemplo, o discurso do entrevistado do EpR ressalta a falta de tempo para se envolver com os resultados do projeto.

> Os gestores não têm o conhecimento disso, de avaliar políticas. Assim, não é uma área ainda que as pessoas não dão muita importância, porque tem várias coisas no seu dia adia que você tem que preocupar, de questões burocráticas, como é que funcionam os processos para gestão desses projetos e tal... Ás vezes ele: "Ah, não, mas meu recurso está ali, eu tenho que fazer é isso, tenho que liberar meu recurso, tenho que fechar esse contrato aqui". Ele não pensa na política. Essa parte de você pensar a política, o que você está querendo implementar, como é que você vai medir isso para ver se é eficiência, isso é eficácia, isso gera impacto. (TÉCNICO EPR2)

> Então o tipo de problema que a gente teve normalmente com manutenção dos dados são muito mais questões operacionais, porque pessoas também na secretaria tinham muitas prioridades, então nem sempre a pessoa está lá disponível para prestar o serviço a qualquer momento, porque as pessoas tinham outros trabalhos. (CONSULTOR IGESP)

Outro ponto relacionado às características da organização diz respeito às fontes de informação utilizadas pelos tomadores de decisões quando confrontados com um problema. Conforme apontado por Cousins e Leithwood (1986), fontes alternativas de informação competem diretamente com os resultados da avaliação para tomada de decisões, sendo o conhecimento prático, crenças, valores e interesses de pessoas envolvidas nos programas fortes fatores adversários da deliberação técnica. O próprio gerente do IGESP confirma o entendimento dos autores, na medida em que relata que, geralmente, as decisões na área de segurança são

tomadas por meio de conhecimento informal, como se fosse perda de tempo utilizar avaliações para uma decisão mais motivada.

> E eles não têm uma lógica assim "Deixa eu te mostrar tudo para você tomar decisão se você vai ou se não vai", eles iam mostrando, eles partem do pressuposto que segurança é fundamental, de que eles sabem o que tem que fazer, (estou falando de polícia) de que eles sabem muito bem o que tem que ser feito, que isso é importante e que o planejamento que tem é para esse ano e se vier mais para o ano que vem tem que ter... Era assim que eles falavam, nesses termos. Eu lembro do gerente da PM falando isso. Falando: "Não tem que ficar planejando, ficar avaliando o tamanho do negócio todo para ver se toma decisão. Isso tem que ser feito e acabou". (GERENTE DO PROJETO IGESP, grifo da autora da dissertação)

Saindo um pouco do domínio da SEDS, observa-se que outros possíveis atores que poderiam ser considerados como aliados da avalição, infelizmente, não foram incluídos. Um dos grupos desconsiderados e que tinham conhecimento da área de políticas públicas de segurança pública foi o Núcleo de Estudos em Segurança Pública (NESP), ligado ao Centro de Estudos em Políticas Públicas da Fundação João Pinheiro. Em linhas gerais, o núcleo trabalha com pesquisas na área de políticas púbicas de segurança e elaboração de boletins e anuários de informações criminais de Minas Gerais, além do treinamento e formação de recursos humanos para as organizações do sistema de justiça criminal brasileiro. De acordo com o entrevistado do EpR, a não participação do NESP foi um contrassenso, já que eles poderiam auxiliar os avaliadores quanto à avaliação, como relatado no trecho seguinte.

> Quando eu cheguei na integração mesmo as coisas já estavam desenvolvidas. O relacionamento do consultor, que é o Consultor junto ao pessoal da Fundação, era muito bom. Então, o Consultor conversava com as pessoas, sentava para participar. E uma coisa que eu achei muito esquisita nessa avaliação é que o pessoal do NESP, eles não participaram ativamente, o que era um contrassenso, já que ali tinham pessoas que integraram com a SEDS, que tiveram cargo. O pessoal do NESP não participou muito não e isso eu acho que era um erro porque eles poderiam contribuir muito mais na hora de desenhar o modelo, de explicar para os pesquisadores, ajudar no que eles estavam medindo. (TÉCNICO EPR2)

Outro grupo que teve a participação reduzida nas avaliações do IGESP foi o Banco Mundial. Embora todas as avaliações tenham incorporado sugestões das pessoas da instituição de desenvolvimento, principalmente no que se refere a métodos de avaliações de impacto, segundo o avaliador, tal ajuda não foi indispensável no IGESP. De acordo com o avaliador, como já dominava as técnicas de avaliação, a dupla (avaliador mais consultor) conseguiu *caminhar com as próprias pernas*.

> Dentro do IGESP a participação do Banco Mundial não foi tão intensa. Eu acho que pelo o que eu vi, ela foi mais bem aproveitada dentro de outros projetos. Lá no nosso projeto a gente já tinha uma metodologia muito bem delineada, muito bem pensada. (AVALIADOR IGESP)

> Veio algumas pessoas do Banco Mundial apresentar métodos de avaliação econométricos de política públicas, mas eu era uma pessoa que já sabia todos eles, então, assim, a interlocução nesse sentido, para mim ela era redundante. Para mim, consultor, para a gente ela era redundante e para outras pessoas que estavam lá não. Então houve mais essa troca de informações natural. A gente conseguiu caminhar com as próprias pernas muito bem, embora tenham dado opiniões no decorrer do processo acerca do nosso projeto. (AVALIADOR IGESP)

Sem a participação da SEDS e sem subsídio de outros especialistas da área, seja o Banco Mundial ou o NESP – o primeiro sendo especialista em avaliação e outro em segurança pública respectivamente –, a avaliação do IGESP foi considerada por alguns entrevistados como uma avaliação isolada e muito focada no método: isolada, na medida em que contou com a participação de apenas dois atores responsáveis pela a avaliação: um avaliador e um consultor[9], sem a interação de demais grupos ao longo de sua implementação e execução; e muito focada no método, dado que os próprios avaliadores tinham perfis mais analíticos, mais ligados a modelos essencialmente quantitativos. Atenta-se para o fato de que mesmo que, as avaliações de impacto utilizem métodos quantitativos por natureza, o foco descomedido pode deixar a pesquisa hermética, o que exclui espontaneamente a participação de outros interessados e torna a pesquisa *distante*, como comentado nas entrevistas. Essa percepção de que a avaliação tenha ocorrido de forma insulada partiu dos depoimentos de outros avaliadores da FJP, como o avaliador VV2 do projeto Viva Vida, e do técnico do EpR que acompanhou a avaliação como demonstram os trechos abaixo.

> O IGESP... Engraçado, talvez pela troca do perfil do coordenador foi o que eu menos participei. Ele trabalhou muito de forma isolada, ele ia comunicando a gente alguma coisa. E lá também aquela questão, a integração era uma coisa que estava começando, a integração das policias, então, eu acho que enfrentou, sim, essa mesma questão: "como avaliar o impacto de uma coisa que está começando?". Eu sei que eles foram fazendo lá alguns ajustes de forma a tentar focar em alguma coisa, vamos dizer, "vamos considerar impacto, se houve reunião, se não houve". Agora, para falar a verdade eu já esqueci um pouco. Eu tive bem menos contato e também eu esqueci. (AVALIADOR VV2)

> Então tinha que pensar; "vamos avaliar alguns projetos estruturadores, algumas ações importantes desse governo, vamos pensar numa metodologia de ponta que estiver melhor, estatística", mas sem conversar com quem estava ali lidando com a

[9] Embora o planejamento da equipe de avaliação do IGESP incluísse a participação de mais um avaliador, o mesmo abandonou o projeto logo no início e preferiu não participar da entrevista deste trabalho com a justificativa de que, infelizmente, não tinha informações suficientes para ajudar.

> prática. O pesquisador ali tem uma ideia: "Ah, o IGESP é isso, a integração é isso, então vamos lá, vamos fazer esse modelinho aqui, eu vou precisar dessa variável". Aí fica muito na manipulação do dado, uma coisa um pouco distante. (TÉCNICO EPR2)

Como contraponto ao isolamento da avaliação do IGESP, consultor e avaliador ressaltam que os dados estavam disponíveis e mais bem organizados que as demais avaliações realizadas na FJP. Essa estruturação da informação permitiu que a modelagem da avaliação fosse realizada de forma mais particular, já que não foi necessário coletar dados ou mesmo organizá-los. Observa-se que a única avaliação em que os próprios avaliadores não criticaram a ausência e desordem dos dados foi a do IGESP. Para o avaliador e consultor o projeto estava maduro suficientemente para ser avaliado, como demonstram os trechos de entrevista.

> Nesse sentido, na verdade a gente tinha a natureza dos dados disponíveis, o período de dados disponível, a informação que a gente tinha sobre o projeto, e etc, na verdade estava bastante estruturado para esse tipo de avaliação. Acho que nesse sentido foi bastante... era mais que suficiente para o que a gente precisava fazer. (CONSULTOR IGESP)

> Já havia o quê? Cinco anos que já havia sido instalado em Belo Horizonte e na maior parte das cidades a gente já tinha aproximadamente três anos com essa experiência. Os dados recebidos por nós eram dados de boa qualidade. A gente optou por modelar dados que a gente acreditava serem bons indicadores de criminalidade, em geral, crimes violentos contra patrimônio e pessoas e homicídios. Por quê? Porque esses são dados que dificilmente existem as chamadas "cifras negras". Então quando você me pergunta: "Olha, será que o projeto estava maduro o suficiente para poder ser avaliado?". Eu digo que sim, não só por uma questão temporal, mas também pela nossa estratégia de modelagem. Quando a gente decide tratar crimes violentos e homicídios, a gente está escolhendo uma variável adequada para esse tipo de coisa. (AVALIADOR IGESP)

Para o consultor e o avaliador que trabalharam na avaliação, o seu único problema refere-se a não aleatoriedade de seu desenho, dada a importância das avaliações aleatórias. De acordo com Khandker, Koolwal e Samad (2010), o método de atribuição aleatória, considerado o padrão-ouro de avaliação de impacto, é o método mais robusto para estimar contrafactuais cuja forma de seleção dos beneficiários acontece por meio de um sorteio para decidir quem receberá os benefícios do programa e quem não receberá, no caso do IGESP, quais regiões integradas receberiam o projeto e quais não. No entanto, como comentado no trecho abaixo, a concepção de uma avaliação aleatória requer um desenho de avaliação concomitante à implementação do projeto, o que nem sempre é possível de acontecer em governos. Para o consultor do IGESP, várias circunstâncias, dentre as quais constrangimentos políticos, não permitem que o desenho aleatório das avaliações se torne comum em governos, argumentos

esses compartilhados pelos entrevistados do EpR e pelo próprio avaliador do IGESP, como retratado abaixo, respectivamente.

> Vamos dizer, hoje em dia o poder de uma avaliação de impacto, é a avaliação concebida antes da implementação do projeto. A avaliação ideal é a que a gente chama de uma avaliação aleatória, que testa um grupo de controle, outro de tratamento e aloca, digamos, a eficácia do projeto aleatoriamente a algumas unidades e vê o que aconteceu. Hoje, seriam, digamos, a concepção perfeita de uma avaliação, mas isso requer, na verdade, um desejo e um desenho de avaliação anterior à implementação do projeto. E isso em várias circunstâncias tem restrições operacionais aí, várias circunstâncias existem restrições políticas mesmo, então é muito comum isso não ser feito. (CONSULTOR IGESP)

> Porque não foi pensado que ia fazer uma avaliação de impacto. Assim, minha opinião é que se quer fazer uma avaliação de impacto você tem que preparar ela antes. Vai medir impacto com relação a quê? Você tem alguns critérios metodológicos para fazer avaliação de impacto e que é complicado. Aí você vai ter que adaptar o programa que está avaliado à metodologia, e eu acho que era o contrário. Então não foi coletado informações antes de implementar as ações. As políticas já estavam sendo implementadas e aí tinha que medir e tinha que ser uma avaliação de impacto. Então, foi complicado porque as pessoas quando foi implementar a política não colheram informações, então você não tinha informação, você tinha que criar tudo a partir do momento que... Então para você ter grupo de controle, ver o antes e o depois é complicado. (TÉCNICO EPR2, grifo da autora da dissertação)

> Isso é muito difícil na maioria dos contextos, no mundo inteiro na verdade, setor de saúde, de educação, em vários países. Então isso, na verdade, não foi feito no IGESP. O que houve no IGESP foi o que a gente chama de uma avaliação *ex-post*, usando a linguagem economicista, uma avaliação posterior, que dada à implementação, dada a forma que a avaliação aconteceu você tenta utilizar o procedimento estatístico que te permita ver se houve um impacto realmente direto do programa ou se era algum outro fator que contribuiu para aquela simultânea implementação do programa. (AVALIADOR IGESP)

Visando somente o aspecto metodológico da avaliação e seu rigor técnico, na visão do consultor, a única limitação da avaliação estaria relacionada à não concepção da avaliação anteriormente ao desenho do projeto realizado de maneira aleatório. E, de fato, conforme visto nos padrões do *Joint Committee* (1994), os parâmetros de rigor ou exatidão determinam se uma avaliação produziu informações boas, relevantes e tecnicamente adequadas e objetivam garantir que a avaliação revele e transmita informações precisas sobre o mérito do programa. Na percepção do avaliador, o pressuposto de que as avaliações deveriam ser formuladas de forma integrada à formulação da política pública deve ser algo perseguido e incentivados pelos governos até que se crie uma cultura.

> O que ficou faltando realmente no caso do IGESP é dar esse último passo. Você começa a pensar numa avaliação antes de implementar um programa. Então eu acho que isso é uma coisa que tem que começar a entrar também, de certa forma, na

> agenda das formações das políticas públicas. Então: "Ah, concebemos um novo programa. Concebemos o IGESP, achamos que isso vai ser bom", na verdade, tem que pensar o seguinte, "Bom, antes de implementar a gente vai querer avaliar, você vai querer fazer funcionar ou não. Então, como eu posso casar a implementação do programa com uma demanda para entender quais são os mecanismos que estão funcionando ou não?" Então, isso não foi feito no caso do IGESP e acho que em algumas dimensões a avaliação poderia ser ainda melhor, certamente, se tivesse sido concebida *a priori*. O fracasso, entre aspas, está aí. A limitação foi essa: avaliações desse tipo, sendo concebida durante o processo de implementação do próprio programa, tendem a ser mais bem sucedidas e mais convincentes. (CONSULTOR IGESP)

Ainda com os constrangimentos metodológicos existentes, o consultor da avaliação reafirma a importância do estudo em termos de consulta e de desenvolvimento de métodos e técnicas para posteriores avaliações de programas desse tipo, como refletido no trecho de entrevista a seguir.

> Tendo em conta uma avaliação como a do IGESP, feita depois que o projeto está implementado com base em dados administrativos, eu acho que foi muito bem-sucedido em termos do tipo de análise que ela poderia fazer. Acho que realmente tem um tipo de informação policial, informação e resultado disso sobre criminalidade, sobre a efetividade da polícia e que, na verdade, não existe muita coisa feita nessa área na leitura internacional. Então acho que nesse sentido foi um sucesso, tá? O grande sucesso foi o resultado da avaliação e o que que ela representou em termos direcionamentos e de jogar luz sobre o efeito de programas desse tipo. Então acho que isso foi muito legal. (CONSULTOR IGESP)

Contudo, observa-se que, ainda que o método constitua um dos atributos essenciais das avaliações, não pode ser considerado como o único aspecto inerente ao seu sucesso ou ao seu fracasso. Quando o pesquisador argumenta que a fragilidade do método, por não ser aleatório, seria a única limitação da avaliação do IGESP, ele desconsidera, portanto, outros aspectos tão essenciais e importantes quanto o rigor técnico como, por exemplo, a cooperação entre os diferentes grupos envolvidos. Na verdade, ao considerar os padrões do *Joint Committee* (1994) o rigor técnico é apenas uma das características destinadas a julgar o nível de aderência da avaliação; a utilidade das informações geradas para os usuários, a sua viabilidade e a adequação ética são alguns dos aspectos essenciais à geração de avaliações informativas, tempestivas e influentes. Os resultados da avaliação do IGESP, sua apresentação, bem como o processamento de informações é o tema da seção seguinte.

8.3 Resultados e processamento de informações

Diferenciando-se das avaliações do projeto Minas Sem Fome e Viva Vida, a avaliação do IGESP apresentou resultados positivos. De acordo com o relatório final da avaliação, os

resultados indicam um efeito do IGESP significativo e quantitativamente relevante sobre a taxa de criminalidade, em particular sobre a taxa de crimes violentos contra o patrimônio, o que apresentou uma redução de 47 a 53% nos locais onde a metodologia IGESP foi implantada (FUNDAÇÃO JOÃO PINHEIRO, 2010c). A evidência também sugere que a implantação do IGESP está associada a um aumento na eficiência da ação policial, refletido no aumento da apreensão de armas por crime cometido e, nos primeiros anos de implantação, a um aumento no número de detenções por crime cometido (FUNDAÇÃO JOÃO PINHEIRO, 2010c). Segundo as sugestões de melhoria apontadas no relatório, como recomendações, além da organização e melhoria dos registros atuais, foram destacadas a possibilidade de utilizar outras fontes de dados, como o DATASUS[10], para verificar os impactos da implementação do IGESP sobre as taxas de homicídio e comparar os resultados com os encontrados utilizando os dados dos registros das polícias, já que o relatório não apontou taxas de homicídio estatisticamente significante para as demais localidades do estado com exceção da Região Metropolitana de Belo Horizonte (RMBH). Ademais, como comentado na entrevista do avaliador, havia um direcionamento claro no relatório para estender a metodologia para os demais municípios do estado, para que esses pudessem experimentar a mesma redução de criminalidade observada na RMBH.

> De fato, a gente faz uma identificação de efeito, a gente mensura esse efeito, e uma vez que esse efeito está medido, a gente, naturalmente, estende esse resultado, essa inferência para os municípios que não têm. Então, o que a gente termina concluindo no relatório é que se essa metodologia foi estendida a mais municípios, o que a gente vai experimentar é uma redução de criminalidade similar àquela que a gente observou. Agora, porque que o Estado não estende isso, aí entra em questões que fugiam ao escopo da nossa avaliação. (AVALIADOR IGESP, grifo da autora da dissertação)

Os primeiros resultados da avaliação do IGESP foram apresentados para o EpR em março de 2010, sendo o relatório de avaliação, em sua versão definitiva e final, entregue em julho do mesmo ano, de acordo com o avaliador da FJP. O entrevistado resume, no trecho abaixo, como aconteceu o decorrer da avaliação do IGESP.

> Nós ficamos dois anos fazendo isso. Durante o ano de 2009, os primeiros oito meses ali foram simplesmente a gente imaginando, delineando como seria essa avaliação, estudando, levantando evidencias de casos similares mundo afora, ver qual seria a melhor estratégia de identificação. Ao final de 2009 começou de fato o período em que a gente começou a atacar essa frente de obtenção das informações e modelagem e identificação de efeitos. Então eu me lembro que havia *deadline* importante, que

[10] Banco de dados do Sistema Único de Saúde.

> era março de 2010. Então, tendo esse *deadline*, a gente entregou isso num primeiro *workshop* interno com a EpR em março. Entre abril e julho a gente atacou ainda mais a obtenção de dados, que eram essenciais para a gente fazer a avaliação, terminamos o nosso procedimento de avaliação, fizemos apresentação dentro de um seminário interno na Fundação, teve um grande seminário lá de todos os projetos, apresentamos isso para o Estado para Resultados e apresentamos isso para SEDS. Primeiro a gente apresentou para a subsecretária. Depois apresentamos para todo o corpo da SEDS. (AVALIADOR IGESP)

Ao tentar reconstituir a cronologia da avaliação do IGESP, a partir da fala do avaliador, observa-se que se destinou muito tempo da pesquisa para a preparação da avaliação, oito meses do ano de 2009 como comenta o entrevistado. Depois disso, já na fase de execução da avaliação em 2010, alguns resultados preliminares foram apresentados para o EpR em março do mesmo ano e, posteriormente, em julho, já de posse dos resultados completos, foram feitas apresentações para o EpR, para os participantes dos outros projetos de avaliação da FJP e para a SEDS, respectivamente.

Os prazos para as entregas, bem como o formato das apresentações sempre foi intermediado pelo EpR, sendo o gestor do projeto e sua equipe os últimos a saber dos resultados, como reitera o avaliador no trecho de entrevista selecionado abaixo. De acordo com a fala do avaliador, apesar de ele mesmo colocar-se à disposição para apresentar os resultados da avaliação para a secretaria de segurança, o interesse maior de que a SEDS tomasse conhecimento das conclusões era do próprio EpR e, devido a isso, os últimos foram os responsáveis pela intermediação. Tanto consultor, como avaliador, demonstram uma posição clara de que após a entrega dos relatórios finais, o trabalho estaria concluído e a forma como se daria o emprego dos dados não seria responsabilidade dos pesquisadores.

> Uma vez que a gente fechou nossos resultados, apresentamos internamente, apresentamos para Estado para Resultados, e ele mesmo, o própria Estado para Resultados, foi atrás da SEDS intermediar uma apresentação. Então, também era o momento deles mostrarem o que eles tinham feito. Era interesse para eles que a SEDS tomasse conhecimento dos resultados de programa de avaliação que eles estavam tutelando. Então, eles é que foram atrás. Nós, da Fundação, não nos envolvemos, mas como o corpo técnico que realizou a avaliação a gente ia apresentar, e é claro, eu tinha todo o interesse de apresentar os resultados do trabalho que eu tinha realizado. (AVALIADOR IGESP)

> A gente teve que fazer entregas diferentes nas diferentes etapas. Então, quer dizer, não foi que a gente fez a avaliação e entregou a avaliação não, na verdade, foi em etapas. Teve o primeiro processo onde a gente fez uma parte que era uma descrição institucional do próprio IGESP, que tinha a questão inicial dos dados. Em algumas áreas que o pessoal deu o *feedback*, aí depois a gente fez uma versão com os resultados, mas ainda preliminar, teve mais um *feedback* e aí a gente entregou uma versão final. Acho que o pessoal ficou bastante satisfeito. Eles acharam que do

> ponto de vista técnico e até da estrutura do relatório entregue eles gostaram muito. Depois disso tudo a gente não teve muita informação sobre o que aconteceu com o IGESP, de como evoluiu. (CONSULTOR IGESP)

No discurso de Patton (2008), os avaliadores, como peças fundamentais da avaliação, ao contrário de somente executar a pesquisa e entregar suas conclusões, teriam um papel mais ativo na promoção e cultivo do uso após a finalização do trabalho do estudo avaliatório. Por não se tratar de uma simples pesquisa científica, mas de uma avaliação para responder questões específicas, o trabalho dos avaliadores deve se estender ao pós-avaliatório, na execução de resumos, como no auxílio na confecção de planos de ação derivados da avaliação e também na implantação desses planos. Para o técnico do EpR, os resultados tinham que ter sido mais debatidos e deveria haver um acompanhamento das recomendações apontadas, como comenta no seguinte trecho de sua entrevista.

> Acho que os resultados tinham que ter sido mais divulgados, debatidos e tinha que ter um acompanhamento dessas recomendações do que foi que aconteceu. Teve uma avaliação de impacto, produziu relatório lá, teve os indicadores e tal, e aí? O que que vai morrer aí? Escutar e ver se isso teve uma mudança dentro do projeto. (TÉCNICO EPR2)

Um ponto que chama a atenção na fala do gerente do projeto diz respeito à dificuldade de interpretar as informações contidas em uma avaliação de impacto, em suas palavras *não é todo órgão e todas as pessoas que tem condição de interpretar*, como resgatado no trecho de entrevista abaixo. Segundo Leviton e Hughes (1981), além da complexidade dos métodos, os jargões e vocabulários específicos utilizados pelos avaliadores constituem um dificultador do entendimento dos resultados das avaliações pelos usuários, uma vez que, para a maioria dos gestores e políticos, a avaliação ainda constitui uma ferramenta incomum.

> Isso é geral. Não é todo órgão e todas as pessoas que tem condição de interpretar. É até um risco, porque se você apresenta resultado disso para um secretário que não tem a capacidade de aprender e aquilo ali, de certa forma, ele interpreta da maneira dele, ele vai tomar decisões assim, extremamente equivocadas. Então, eu acho que assim, a avaliação tem que ter aderência com a lógica gerencial, você tem que ter um corpo gerencial que saiba o que está fazendo com aquilo e que não seja no órgão. Então, é uma avaliação externa e esse corpo gerencial desse órgão externo sabe o que aquilo significa, sabe o que que ele está fazendo, porque senão é complicado. E é isso, tem uma questão de maturidade gerencial que ela é alcançada com o tempo, mas tem que fazer. Acho que tem que fazer a partir de um lócus assim, que saiba o que está fazendo com propriedade, que tem que está em conjunto, tem que criar essa capacidade no órgão. (GERENTE DO PROJETO IGESP, grifo da autora da dissertação)

Na percepção de alguns entrevistados, a natureza dos resultados da avaliação do IGESP cooperou com o processamento de informações para a SEDS, já que os resultados da avaliação quanto ao impacto do projeto na redução da criminalidade foram positivos. Nas palavras do entrevistado, *estavam menos preocupados porque os resultados, de certa forma, eram positivos, se fossem negativos aí seria mais complicado*. E, ao mesmo tempo em que a natureza dos resultados tornou o processo menos conflituoso, também ofereceu menos oportunidades para discussão e reflexão a partir da avaliação. Como explicam os trechos de entrevistas abaixo selecionados, embora reconheçam a importância do estudo para identificar a efetividade do IGESP, em termos de reconstrução e melhoramento da política assumem que talvez o instrumento de mensuração apenas do impacto do projeto não tenha sido o mais apropriado para responder as perguntas sobre a operação do projeto.

> Na parte da defesa não teve muito interesse. Apresentou, acharam bom porque o resultado era positivo. (TÉCNICO EPR2)

> Eu disse "ok", isso é ótimo, isso fortalece, isso legitima mais, mas a questão que a gente lidava era... Era exatamente assim: "Gente, como é que opera isso? Por quê? Porque que ele gerou isso? porque que o IGESP gerou? O que do IGESP que gerou legitimidade?". Isso não é respondido. Então assim, é bom, isso te assegura que você está no caminho, que é isso mesmo, mas não esgota o que você precisa saber no dia a dia, que é "como é que eu faço esse método melhorar? (GERENTE DO PROJETO IGESP)

Nesse sentido, atenta-se para as diferenças existentes entre as avaliações de impacto que aludem aos efeitos do projeto a partir de metodologias essencialmente quantitativas, e as avaliações de processos que representam chances ao monitoramento dos processos responsáveis pela produção do efeito esperado da política (PEREZ, 2001). Como apresenta Perez (2001), a pesquisa sobre os processos de um programa não se restringe apenas a diagnosticar, mas a corrigir os cursos das ações no momento em que os programas estão sendo executados. Na percepção do gerente do projeto, o gestor deixa indícios de que, caso a avaliação fosse qualitativa e representativa das questões de implementação da política de gestão integrada, talvez seria mais bem aproveitada, uma vez que aponta que a avaliação de impacto ficou restrita. Alguns atores, como Leviton e Boruch (1983), corroboram a percepção do gestor na medida em que argumentam que estudos avaliatórios focados em desenho ou processos seriam mais bem utilizados que estudos focados em resultados dos programas.

> A de impacto eu acho que ela ficou muito restrita, muito. Engraçado, né? Por quê? Primeiro porque talvez ela não seja qualitativa, porque o negócio dela é muito mais o método. E para quem, na verdade, lida com o programa, para falar a verdade, o

> método, a modelagem estatística você não tem capacidade de criticar aquilo, de saber se aquilo está certo. Eu não tenho. Você vai na base da confiança: "Olha, eu acho que esse povo sabe o que está fazendo, porque esse 'trem' é muito complicado". (GERENTE DO PROJETO IGESP)

Tentando responder ao problema de pesquisa evidenciado neste trabalho, qual seja o questionamento de como os diferentes *stakeholders* envolvidos nos processos de avaliação percebem a utilidade dos resultados das avaliações, desenvolvem-se, na seção seguinte, os diferentes entendimentos sobre o uso da avaliacao do IGESP na prática do projeto.

8.4 Utilização da avaliação do ponto de vista dos *stakeholders*

Cousins e Leithwood (1986) afirmaram em seu estudo que descobertas esperadas e positivas estariam mais relacionadas à aceitação e à utilização por partes das equipes dos programas e que resultados surpreendentes e que exigem mudanças significativas no projeto, por seu turno, estariam fadados ao não uso. A dinâmica de utilização dos resultados da avaliação do IGESP, no entanto, ainda que apresentasse resultados positivos, não foi mais tranquila que as demais avaliações que apresentaram resultados negativos. A mesma divergência de opinião entre os diferentes atores participantes pode ser observada, o que dificulta um consenso sobre a utilidade do estudo avaliatório.

A percepção do avaliador é de desconhecimento quanto ao uso da avaliação, uma vez que, em sua visão, o que seria feito após a conclusão da avaliação não é entendido como escopo de trabalho de avaliadores. Como já comentado anteriormente, para o avaliador, a função de avaliadores é demonstrar a *verdade*, se o projeto tem efetividade ou não, tarefa essa que foi cumprida na opinião do avaliador do IGESP. O entrevistado apenas supõe que, dada a natureza dos resultados, por serem positivos, crê que eles tenham sido, de alguma forma, vinculados na mídia já que, na sua percepção, é esperado que resultados positivos sejam noticiados e que os negativos sejam suprimidos.

> Nosso cliente era o Estado, então, a gente chegava e entregava isso para ele. O que ele ia fazer com essas informações já não era mais da nossa alçada, mas ele tinha interesse em saber a verdade, se aquilo era mero desperdício de recurso ou se aquilo era relevante. Eu creio que isso tenha sido vinculado de alguma forma na mídia, que isso tenha sido usado de alguma forma, claro, porque, poxa, dado que os resultados são bons não tem como não divulgar. Os governos fazem isso, é esperado que eles façam isso – eles vão divulgar seus bons resultados e não vão divulgar os resultados que não são bons. (AVALIADOR FJP, grifo da autora da dissertação)

Embora o avaliador tenha demonstrado insegurança em seu relato quanto ao uso da avaliação, na percepção do entrevistado da coordenação da FJP, a certeza é de que o uso do relatório do IGESP não aconteceu, ainda que tenha demonstrado a efetividade do projeto. O entrevistado explica que a rotatividade da alta direção da SEDS, da própria equipe do projeto de integração e do avaliador que trabalhou na avaliação prejudicou, não só a utilização, mas também a sedimentação de conhecimento quanto ao tema de avaliações. O uso processual, relacionado ao aprendizado gerado pela realização de avaliações, comum nos demais projetos avaliados da FJP, não aconteceu, tanto por ser uma avaliação construída de forma isolada e rápida, quanto pela saída de todas as pessoas envolvidas. Para o entrevistado, da avaliação do IGESP *não ficou nada*, como comenta nos trechos abaixo relacionados.

> Eu acho que mostrou que foi uma boa política. Eu acho que ninguém fez uso do relatório. Ninguém se importou. Mas também nós apresentamos o relatório para o secretário e logo depois ele saiu, entrou nova pessoa, e foi no final do ano. As pessoas para quem a gente apresentou o relatório no final de 2010 saíram logo depois. O gerente do IGESP saiu, a secretaria de segurança, o secretário mudou, então se perdeu nessa mudança de governo. Mas eu acho que mostrou bem que a coisa funcionou. As pessoas entraram e saíram o tempo todo. No fim ficou o avaliador da FJP e o consultor. Concluíram o relatório, ficou muito bom. Acho que é um relatório excelente. Mas foi isso. O avaliador depois saiu também, então, não ficou nada disso. Da experiência da avaliação da política de segurança não ficou nada. (COORDENAÇÃO FJP)

Seguindo a percepção da coordenação da FJP, os entrevistados do EpR concordam que a avaliação do IGESP não tenha desencadeado mudança e nem criado capacidade de avaliação na SEDS. Mesmo que a avaliação tenha contribuído para ser mais um fator para reforçar e justificar a necessidade da política de integração, para o entrevistado da coordenação do EpR, o esforço realizado é análogo a qualquer pesquisa acadêmica de caráter individual, em que se entrega um relatório de conclusão final e o trabalho é considerado encerrado. Para constituir, de fato, uma avaliação, seria preciso que a avaliação do IGESP fosse vista como um processo intermitente, de construir a avaliação como um processo continuado de melhoria ininterrupta das ações do projeto, como argumenta o entrevistado no trecho selecionado abaixo.

> A parte da defesa foi o mais comum porque foi muito rapidinho, foi muito feito mo laboratório, foi apresentado e beleza, mas não impactou tanto na mudança de rotina e não foi uma coisa continuada, ou seja, não se pensou no processo. Porque, assim, a avaliação não é um projeto, a avaliação é um processo, e esse é o grande drama. O IGESP, por exemplo, foi um projeto, não foi um processo, então fez, entregou o relatório e acabou. É uma dissertação de mestrado, é um *paper*, assim, não é um processo que fica na secretaria. A secretaria entende a importância daquilo e de fazer aquilo de modo continuo e buscando sempre aprimorar e tal, e não desenvolveu essa capacidade na SEDS. Eu acho que ela foi muito pontual. Pode ter mudado alguma coisa, mas acabou ali. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

De acordo com o técnico do EpR que participou da avaliação, o instrumento avaliatório deveria ser uma ferramenta para o gestor dos programas, para que eles pudessem repensar as políticas de forma a aprimorá-las continuamente, o que não ocorreu no caso da avaliação do IGESP. Para o entrevistado, a avaliação *não foi um conhecimento que foi absorvido pela SEDS* e exemplifica que muitas pessoas que deveriam ser interessadas com o trabalho se referem a eles como *uma pesquisa aí do Estado para Resultados*. Não houve, portanto, um sentimento de apropriação das informações geradas, como se depreende por meio dos trechos.

> Acho que a avaliação, ela tem que ser importante não só para os órgãos mais estratégicos, especializados em planejamento, mas ela tinha que ser um instrumento mais próximo do gestor das políticas, porque a tendência são esses gestores, e que tivesse um acompanhamento durante todo o processo, desde pensar a política, antes de pensar, para poder conseguir fazer um marco lógico daquela política, desenhar adequadamente, ver as variáveis importantes para você mensurar, para fazer algum tipo de controle. Acho que isso ajudaria muito. Muitas pessoas que você perguntava, da Defesa, pessoas que são parte interessada, nem conhecem essa pesquisa. Vai saber: "Ah, tem uma pesquisa aí do Estado para Resultados!", mas não foi um conhecimento que foi absorvido pela SEDS. (TÉCNICO EPR2, grifo da autora da dissertação)

Já o gerente do projeto, diferentemente das percepções anteriores, consegue discernir pontos positivos da avaliação. Na visão do gerente da política de integração, a avaliação foi um importante instrumento para confirmar a orientação acertada das ações implementadas e, assim, resguardar o projeto de prováveis interferências políticas. Fazendo uso da teoria explicitada por Patton (2008), esse caso remete ao uso persuasivo, em que os resultados das avaliações são utilizados para dar suporte aos debates políticos e decisões específicas, como, por exemplo, discussões sobre orçamento público. Nesse sentido, o gestor explica que a função da avaliação foi de blindar o projeto de possíveis ataques políticos, tanto que, no momento em que o estudo foi apresentado, ainda que prevalecesse o contexto de restrição orçamentária, houve uma confirmação de que a política seria expandida para todo o estado de Minas. E exemplifica a situação: sempre carrega a pesquisa *debaixo do braço* e quando alguém questiona se o projeto deve continuar ou não, mostra a pesquisa como um indício objetivo da efetividade do projeto, como se a avaliação fosse a *arma* contra as interferências e restrições governamentais impostas.

> Olha, poderia ser assim, eu acho que ela veio confirmar e fortalecer, porque tudo na integração está sempre em cheque. Todo o momento as pessoas estão assim "Ah, acho que talvez não valha mais a pena reunir. Será que esse IGESP está funcionando?". É sempre polêmico, sempre difícil. Então, eu acho que assim,

> legitimidade nesse meio é fundamental, então, eu acho que fortaleceu. O IGESP fica protegido, do ponto de vista de alguns questionamentos em função do embasamento de uma pesquisa, isso é claro. Tanto que, quando a pesquisa chega, o resultado dela, o IGESP ainda não estava em todo o estado. E na hora que ela chega, num contexto de restrição orçamentária, há uma confirmação de que vai expandir para o resto mesmo não tendo recurso. Ou seja, há uma decisão de continuar a expansão e de priorizar os recursos necessários para a expansão do IGESP num contexto de restrição orçamentária. Eu acho que é isso. É o fortalecimento do reconhecimento da legitimidade do projeto. Assim, já era uma expectativa, a gente já esperava porque ele já era visto como prioridade, mas confirmar eu acho que é uma coisa... Mesmo porque quando a pesquisa chega, o resultado, você já está num momento de transição política. Então, assim, o Secretário está saindo já, logo depois já é outro, então você sempre fica com aquele negócio debaixo do braço. Na hora que o povo fala assim:
> —Vamos parar esse IGESP.
> Você fala:
> —Não, aqui ó, tem grande efetividade.
> Então, assim, você usa o que você tem. Você fala assim:
> —Não, tem uma pesquisa mostrando que mesmo sem os investimentos...
> E por aí vai. (GERENTE DO PROJETO IGESP, grifo da autora da dissertação)

No entanto, ainda que a avaliação tenha confirmado a decisão de expandir o projeto para as demais cidades do estado, o gerente demonstra que a essa deliberação foi precipitada. O projeto foi ampliado, mas não foi mantido com o mesmo nível de investimento anterior à avaliação dado o contexto de restrição orçamentária vivido pelo governo naquele período, o que, na visão do gerente, contrariava o desenho da política. Em diferentes pontos da entrevista o gerente do projeto comenta como o uso atropelado dos resultados da avaliação pode ter contribuído para que uma decisão equivocada fosse tomada.

> Eu não percebia isso à época, mas, hoje, vendo tudo, eu acho que essa pesquisa veio confirmar uma expectativa, ela confirmou uma expectativa que a era de expandir, e com base nela essa decisão foi tomada, ela foi confirmada. Quer dizer, já se queria, a pesquisa chega, "Então vamos. Vale a pena". Porque ela é efetiva. A metodologia sozinha funciona, então vamos. E foi. E não foi bom. Não foi bom. Vamos pensar assim, talvez esse negócio empolgue demais e tenha sim contribuído para que uma decisão equivocada fosse tomada. Na hora que essa pesquisa vem em 2010, o IGESP tem efetividade dele, bacana, então está no caminho certo, mas ao mesmo tempo o dinheiro acaba. "Não, vamos continuar funcionando IGESP", sem investimento nenhum, nenhum. Então, assim, um pouco as coisas, elas são meio que usadas às vezes para tomar decisões que, de certa forma, às vezes, até contrariam a política. Eu acho que isso era uma realidade, não tinha recurso mesmo. E aí vamos pensar, se era um gestor, você fala: "Ah, vamos então manter esse método aí que alguma coisa ele vai dar de redução mesmo sem a gente está continuando". E todos os outros municípios que receberam depois o IGESP, eles, coincidentemente, já não tiveram mais os mesmos investimentos. E a gente assim: "Nossa, a policia quer o projeto. quer expandir o projeto. Bacana". Vem isso e a gente fala? "É isso aí!". Algum grau de contribuição ele tem sozinho, então vamos embora. Está sem dinheiro, tem efetividade, então vale a pena apostar. Só que aí a gente fala: Foi decisão errada? Foi. Hoje eu falo que foi. A decisão que eu ajudei a tomar foi equivocada, não foi a melhor decisão. Mas também foi por causa da pesquisa? Não. Acho que a pesquisa foi um elemento. (GERENTE DO PROJETO IGESP, grifo da autora da dissertação)

Entram em cena, assim, as diferentes possibilidades oferecidas por uma avaliação: de tanto melhorar uma política, recomendando modificações na formulação, na implementação ou nos resultados; como prejudicá-la caso não seja esclarecida e discutida até o seu esgotamento, como reflete o gerente do projeto. Atenta-se, portanto, para a importância do momento do pós-avaliatório, período em que o conteúdo, a linguagem, o contexto do programa avaliado e os valores dos usuários devem ser discutidos e negociados, para que a implantação das recomendações seja algo consensual e adequado às circunstâncias vivenciadas pelo projeto na etapa de entrega da avaliação, como aborda Patton (2008). Muitas vezes, os próprios gestores não sabem como aproveitar o conteúdo das avaliações por ausência de conhecimento dos usuários quanto ao tema, constituindo, assim, uma não utilização de forma não intencional, como abordado por Alkin e Coyle (1988). Nesse sentido, a descontinuidade da avaliação, somada ao desconhecimento dos gestores de como conduzir as descobertas de um estudo, podem levar ao fracasso do investimento realizado, como argumenta o gerente do projeto.

> E aí vem uma coisa talvez final, é também a descontinuidade da avaliação. Isso é muito sério, você avalia, toma determinadas decisões com base na avaliação, porque toda decisão... a avaliação, ela te dá determinados parâmetro, ela te faz enxergar a realidade sobre uma determinada dimensão, você toma decisões. Nós estamos lidando num ambiente que é muito complexo, que você não consegue calcular outras coisas não previstas, então você precisava que isso fosse mais sistemático. Você tem que saber usar. Pesquisa é uma coisa boa, ou seja, ela é uma ferramenta, e ferramenta você faz coisa boa e coisa ruim nela. Então, assim, as pessoas tem que esta preparadas, tem que se ter um pouco de conhecimento para você absorver o que que aquilo quer dizer, porque senão precipita. (GERENTE DO PROJETO IGESP, grifo da autora da dissertação)

Por mais bem intencionadas que sejam, ainda que tenham como objetivo apoiar os processos gerencias a partir da geração de informações, nota-se, claramente, na fala do gerente do projeto, que não valerão de nada se os próprios gerentes não souberem como utilizá-la. Sobre esse fato, a falta de familiaridade por parte dos usuários das avaliações de como contratar, conduzir, revisar e, posteriormente, usar as avaliações limita o seu potencial de utilidade, como afirmam Bamberger, Rugh e Mabry (2011). Para o gerente do projeto, os gestores, na verdade, não têm *expertise* para encomendar pesquisas que trarão as respostas certas aos seus questionamentos e nem para interpretar as informações quando os estudos são concluídos, já que os próprios gestores não sabem o que querem quando se trata de avaliações, como reflete o gerente nos trechos abaixo.

> Você não nasce sabendo. Muita coisa de pesquisa que a gente fez, das outras, elas encomendou e foi aquilo que o pesquisador sabia fazer, não é necessariamente aquilo que você precisa. Isso eu percebo claramente hoje. Porque a gente ainda não

> sabe exatamente o que precisa, não sabe. Do ponto de vista da avaliação não sabe que perguntas precisam ser respondidas. Acho que em relação ao IGESP a gente sabia mais ou menos assim, pela trajetória dos envolvidos, porque foi o CRISP que fez, estava no início da implementação, porque os pesquisadores que fizeram tinha uma trajetória, porque a gente tinha. Então, para mim ela até se aproxima, a do IGESP, qualitativa assim, ela se aproxima muito daquilo das perguntas que a gente precisava que fossem respondidas minimamente. Depois que ela terminou a gente falou: "Nó, acho que a gente tinha que saber agora mais isso!". Ela não se esgota. É assim: "Mas e isso? E aquilo?". Essa estrutura tinha que ser contínua, porque o método está rodando, foram continuados, então, tinha que ser continuado. E a de impacto também. Agora, várias outras, estou me referindo a outras, é uma coisa de avaliação geral, e a gente fica perguntando assim: "De onde surgiu isso?" Quem fez essa pergunta? O que que isso responde? Tem isso. Então, assim, você vai falar para um decisor: "O problema é da pesquisa?". Não, não é da pesquisa. Eu acho que a pesquisa foi extremamente útil, mas será que a gente está sempre preparado para interpretar tudo que a pesquisa quer dizer? Será que é da mesma maneira? (GERENTE DO PROJETO IGESP, grifo da autora da dissertação)

De fato, as avaliações só irão constituir um instrumento válido quando conduzidas de maneira adequada e com objetivo bem definido. Escolher uma ferramenta de avaliação, ou uma combinação de ferramentas, depende, primeiramente, da definição de qual informação se pretende buscar e das possibilidades e recursos existentes para se conduzir um estudo de qualidade. O tempo de duração, o foco e o nível de detalhes de uma avaliação devem ser determinados pelas necessidades de informação e tipos de decisão dos próprios *stakeholders* (BAMBERGER; RUGH; MABRY, 2011). A ausência de orientação sobre avaliações de programas governamentais, suas metodologias, como acompanhá-las e utilizá-las, somada à ausência de formação dos gerentes dos programas governamentais e formuladores de políticas, constituem limitações ao desenvolvimento de trabalhos de sucesso nesse âmbito, dificultando o entendimento e desenvolvimento da área. De acordo com o coordenação da FJP, as dificuldades presentes nos processos de avaliação desenvolvidos são também oriundos da falta de experiência na área, um processo em que estavam *aprendendo e fazendo* ao mesmo tempo, como exemplificado no relato de entrevista abaixo.

> Acho que o problema foi mais nosso, de condução da coisa mesmo e isso tem a ver com a falta de experiência. Era a primeira vez que estava se fazendo isso. Era um processo novo, estávamos aprendendo essa coisa e fazendo. Então, isso foi um aspecto negligenciado e eu acho que contribuiu também para que a coisa... Não houve essa interação contínua, intensa nas equipes dos projetos. Então, se apareceu lá no começo da coisa, se voltou dois anos depois com o relatório. (COORDENAÇÃO FJP, grifo da autora da dissertação)

Como fechamento do caso da avaliação do IGESP, torna-se importante ressaltar a natureza conflituosa do projeto por si só, marcada pela superação de uma cultura de trabalho individual para difundir uma crença de que a unificação das polícias ou a integração de suas ações

levaria a uma maior eficiência da ação policial. Em relação à implementação da avaliação do projeto, ressaltam-se, primeiro, o distanciamento entre avaliadores que conduziram o processo e os gestores e equipe que trabalhava no projeto, sem considerar as necessidades de informação por parte desses; e, segundo, o foco demasiado em questões metodológicas em detrimento dos arranjos institucionais existentes. Quanto à utilização dos resultados, observou-se um uso indevido da avaliação ocasionado, em parte, pela ausência de discussão das conclusões até o seu total esgotamento. Como elucidado pelo próprio gerente do projeto, por mais bem intencionadas que sejam, ainda que tenham como objetivo apoiar os processos gerencias a partir da geração de informações, nota-se que as avaliações não valerão de nada se os próprios gestores não souberem como utilizá-la e mais, não tiverem incentivos e auxílios dos avaliadores com o período do pós-avaliatório. Observa-se também, que, diferentemente de outros projetos avaliados, não foi desenvolvida no órgão gestor do programa de capacidade de avaliação.

9 AVALIAÇÃO DO PROJETO MINAS SEM FOME

A avaliação do projeto estruturador Minas Sem Fome constitui uma das avaliações do pacote organizado pela FJP e EpR em conjunto com o Banco Mundial e, portanto, contou com a participação de consultor contratado e avaliadores da FJP. Para a pesquisa empírica deste trabalho, contou-se com a participação de seis entrevistados específicos da avaliação do projeto, o que incluí: gerente do projeto MSF à época em que foi negociada a avaliação, um técnico da EMATER que atuava na execução de uma das ações do projeto e quatro avaliadores da FJP que participaram em diferentes etapas do processo de avaliação. Também incluem-se percepções de entrevistados da coordenação do EpR, do técnico do EpR que acompanhou o projeto e da coordenação da FJP. Observa-se que a entrevista com quatro avaliadores da avaliação do MSF foi necessária devido à alta rotatividade entre os avaliadores desse projeto, já que um foi responsável por iniciar a avaliação, outro pela execução e outros ainda participaram do fechamento e apresentação dos resultados.

O capítulo está organizado em quatro seções que remontam o histórico da avaliação do MSF, a partir da percepção dos entrevistados, desde a concepção da avaliação e expectativa inicial, até a sua conclusão, com alusão ao tipo de utilização ocorrido. Na primeira seção apresenta-se o contexto em que a avaliação foi inserida, seguido pela implementação e envolvimento dos atores na execução da avaliação. A terceira e quarta seções tratam, respectivamente, dos resultados e da utilização da avaliação.

9.1 Contexto geral da avaliação

O enfrentamento do problema da fome, em caráter emergencial, representa uma estratégia basilar de redução da pobreza relacionada ao baixo desenvolvimento da agricultura familiar e à situação de vulnerabilidade social das famílias e comunidades, inclusive, nas regiões periféricas ao meio urbano. Para fazer frente a esse contexto, o projeto estruturador Minas sem Fome, criado pelo governo de Minas Gerais, em 2004, e adjacente à implementação do Programa Fome Zero, lançado pelo Governo Federal, em 2003, tem como objetivo geral "[...] promover a segurança alimentar e nutricional de famílias mineiras em condições de vulnerabilidade social, principalmente no meio rural" (FUNDAÇÃO JOÃO PINHEIRO, 2010b). Em linhas gerais, na tentativa de alcançar seu objetivo, o projeto propõe ações voltadas para projetos coletivos de produção, processamento de alimentos e geração de renda:

fomento a projetos de produção de alimentos para subsistência e geração de renda; implementação de projetos de agregação de valor à produção nas comunidades; implementação de projetos de infraestrutura básica e apoio e capacitação de jovens e do público beneficiado com essas ações.

O público-alvo do MSF é constituído pela população rural composta por agricultores familiares em situação de maior vulnerabilidade social, organizados em associações comunitárias ou de agricultores familiares legalmente constituídas. Sendo que escolas, creches, asilos e entidades assistenciais também podem ser beneficiários do programa. A execução do Projeto é de responsabilidade da Empresa de Assistência Técnica e Extensão Rural de Minas Gerais (EMATER), que atua sob a coordenação das Secretarias de Estado da Agricultura, Pecuária e Abastecimento (SEAPA). A EMATER é a gestora do projeto, atuando no levantamento das demandas, no cadastramento dos projetos apresentados por meio das associações rurais e prefeituras, na articulação de parcerias para a distribuição dos insumos agrícolas, e na implantação dos projetos. Trabalha também na capacitação e assistência técnica dos beneficiários.

A avaliação do MSF foi implementada em 2008 como uma das avaliações planejadas no contrato do Banco Mundial com o governo mineiro, esforço já descrito no capítulo anterior. No entanto, na visão dos gestores da EMATER e avaliadores do projeto, o contexto de avaliação do projeto não estava tão claro. Os entrevistados não souberam, ao certo, os motivos que levaram os dirigentes do governo a selecionar o MSF para passar por uma avaliação de impacto. A justificativa possível, na percepção de avaliadores e gestores do projeto na EMATER, é que o MSF seria avaliado por dois motivos. O primeiro se refere ao fato de ele fazer parte do conjunto de projetos estruturadores; e o segundo se deve aos resultados positivos no monitoramento que, até o momento, não tinham sido comprovados. De acordo com os entrevistados,

> até aonde a gente sabe, ele tinha necessidade também de avaliar alguns projetos estruturadores. Como o Minas Sem Fome é um programa de razoável alcance aí, acho que ele foi avaliado pela importância dele, por ser um dos estruturadores. (GERENTE DO PROJETO MSF)

> Eu imagino que ele foi avaliado porque ele era um dos projetos estruturadores. (AVALIADOR MSF1)

> Eu acho que eles escolheram projetos que eles tinham uma visão positiva no monitoramento. (AVALIADOR MSF4)

De início, é possível observar o desconhecimento das partes envolvidas, avaliadores e gestores do projeto principalmente, sobre a possível explicação e os motivos que levaram o alto escalão do governo a escolher o MSF com um dos projetos constituintes da agenda de avaliações de impacto. Verificam-se, a partir da fala de avaliadores e do gerente do projeto, possíveis palpites sobre a escolha do MSF, quase como um *chute*, mas nenhuma afirmação muito confiante. No entanto, ainda que os motivos da avaliação não fossem claros para avaliadores e usuários da mesma, de acordo com o gestor do projeto, realizar uma avaliação externa do projeto era um anseio da EMATER, vontade esta que pôde ser confirmada pela tentativa anterior da empresa na contratação da Fundação de Desenvolvimento da Pesquisa (FUNDEP) para apostar em avaliações realizadas com recursos própria da instituição.

> Do ponto de vista da gestão do programa, acho que é muito bom que a gente fosse avaliado. Havia até uma demanda de avaliação do programa. Nós tentamos contratar a FUNDEP para fazer essa avaliação com recursos próprios, mas não gostaram do ponto de vista jurídico. Nós queríamos contratar a FUNDEP por inegibilidade, o jurídico não concordou de ficar sem licitação. Nós ficamos em dúvida se valeria à pena abrir uma licitação assim, aberta. Abrir um pregão para contratar podia aparecer qualquer empresa aí. Logo, a SEPLAG tomou essa iniciativa, nós achamos ótimo. (GERENTE DO PROJETO MSF)

O grau de importância da avaliação para os públicos que a receberam, segundo Cousins e Leithwood (1986), constitui uma variável significativa para o posterior aproveitamento das avaliações. Para os autores, estudos demonstram que o comprometimento e o envolvimento com a avaliação promovem uma melhor utilização dos resultados das avaliações (COUSINS; LEITHWOOD, 1986). De acordo a percepção dos avaliadores da Fundação João Pinheiro que participaram da avaliação do MSF, a boa receptividade dos gestores do projeto com a avaliação constituiu um diferencial da avaliação do MSF, fato este que não pode ser observado de forma unânime em todas as avaliações analisadas nesta dissertação.

> Era uma instituição que realmente queria ser avaliada, então eu acho que isso faz toda diferença, o que eu acho que acabou não acontecendo com outras avaliações e que na realidade as instituições não queriam ser avaliadas, e aí você tem toda uma resistência que dificulta o processo. Isso era uma coisa que eles sempre diziam, que a EMATER nunca tinha sido avaliada e eles queriam muito que fosse avaliada. E em momento algum houve qualquer tipo de interferência por parte da EMATER. Eles nunca se opuseram a fornecer qualquer informação ou tentar direcionar qualquer tipo. Eles realmente queriam ser avaliados e isso ajudou muito no processo. (AVALIADOR MSF3, grifo da autora da dissertação)

> Eles nunca tinham sido avaliados externamente, então, para eles, isso era interessante, apontar quais que eram as falhas, as eficiências para que eles pudessem melhorar. Sempre nessa postura. Foi muito bom trabalhar com a EMATER. Gostei muito. Eles sempre muito dispostos a nos receber. (AVALIADOR MSF3)

De acordo com Shulha e Cousins (1997), a estrutura e os processos organizacionais nos quais as avaliações são aplicadas são um fator decisivo para a inserção de mudanças em qualquer organização. Cousisn e Leithwood (1986) completam que as características pessoais dos coordenadores dos projetos avaliados, como liderança, são fatores que impactam positivamente o uso das avaliações. Na visão dos organizadores do processo de avaliação, a boa receptividade quanto às avaliações foi uma combinação entre as características organizacionais do órgão e o perfil de liderança do projeto. Para o entrevistado EpR1, a EMATER se destacou quanto à receptividade por constituir um órgão menos cobrado pela unidade central de planejamento do governo que as demais secretarias avaliadas e, portanto, mais carente pela atenção do governo central e por exercer uma liderança esclarecida quanto à importância das avaliações, como destacado no primeiro trecho abaixo. Por constituir, contudo, uma área menos requerida pelo órgão de planejamento, apresenta-se como uma área mais desorganizada – como comenta o entrevistado da coordenação do EpR2 no segundo trecho de entrevista, com problemas relativos à disposição dos dados que comprometem a própria capacidade de realização das avaliação. A percepção da coordenação do EpR é clara nesse sentido, como demonstram os trechos abaixo.

> Onde mais teve abertura, por exemplo, no Minas Sem Fome. Por quê? Porque onde a vaidade era menor, onde o nível de atenção era menor. Por quê? Porque ele já tinha sido um projeto muito bom em termo de ser prioritário, não era mais tanto assim porque já tinha dado um 'caldo'. A EMATER é uma instituição que tem uma carência de querer ser importante e foi super bem recebida. E tinha um líder [gerente do projeto] que era um tanto quanto esclarecido no sentido dessa abertura, então, assim, foi muito mais rápido. Qual o problema? Tem menor capacidade, tinha dados mais problemáticos e tal e tinha problema com sua equipe, Fundação e execução dela. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

> O Minas Sem Fome que eu não lembro um pouco, mas é uma área menos organizada mesmo; desenvolvimento social. Tanto que a gente pegou um programa que estava na EMATER. Não é uma área muito organizada. (COORDENAÇÃO EPR2)

As expectativas dos gestores do projeto, quando foram comunicados que haveria uma avaliação do MSF, eram duas: que a avaliação do MSF pudesse recomendar melhorias no programa, o que permitiria a correção de rumos do projeto; e que comprovasse a existência de efeitos positivos do projeto, nas palavras do gerente do projeto, *que pudesse mostrar para o*

nosso próprio corpo técnico a eficácia, a eficiência e a utilidade do Minas Sem Fome como uma política pública do governo, como destacado no trecho de entrevista abaixo. Nesse sentido, ainda que o gerente do projeto tenha recebido bem a avaliação, parece por demais óbvio seu entendimento por um caminho unilateral da avaliação, qual seja o de demonstrar o impacto positivo do MSF.

> Nós tínhamos no fundo de gestão duas expectativas. Uma de que essa avaliação nos permitisse de corrigir rumos. Fosse importante para detectar pontos de melhoria no programa. Outro, é que ela pudesse nos dar indicativos, que pudesse mostrar para o nosso próprio corpo técnico a eficácia, a eficiência e a utilidade do Minas Sem Fome como uma política pública do governo. Porque qualquer política pública, qualquer ação é o seguinte: "isso está sendo investido mesmo? Isso é bom mesmo? Está valendo a pena?". Quando a gente faz internamente pode ter desconfianças. Quando alguém avalia com isenção, tem uma ideia mais pragmática quanto à verdade ou não. Então a expectativa nossa é isso, uma avaliação e confirmar o que a gente já achava; que o programa tinha utilidade, eficácia e conseguia alcançar resultados bons na área dele. (GERENTE DO PROJETO MSF, grifo da autora da dissertação)

> É que estaria sendo feita essa proposta de pesquisa aí. Expectativa nossa era realmente de que apontasse, vamos dizer assim, o resultado das ações, né? Toda avaliação que você faz o objetivo é justamente analisar aquilo que está sendo positivo e aquilo que pode ser melhorado. (EQUIPE DO PROJETO MSF1)

O gerente do projeto parece desconsiderar, portanto, as várias possibilidades de uma avaliação. Ao contrário de constituir uma possibilidade para exaltar características positivas do projeto, o valor real de uma avaliação destina-se a identificar, obter e proporcionar de maneira válida e confiável informações relevantes para apoiar um juízo sobre o mérito de um programa, independentemente do tipo de mérito, como demonstram Aguilar e Ander-Egg (1995). Além disso, avaliadores que tendem a produzir as informações que os clientes querem, estariam corrompendo um princípio ético das avaliações, como atenta Guba (1975). Entretanto, no caso do MSF, a esperança por resultados positivos por parte dos usuários, somada à sua boa receptividade, gerava certa apreensão por parte dos avaliadores, como sugerem as entrevistas a seguir.

> Sinceramente, eu pensava isso todo o tempo "E aí, e se não for detectado impacto? Se a leitura for negativa?" E eu acho que havia uma preocupação geral em relação a isso. Tanto que havia certo cuidado em relação a publicidade, estava dito que não seria de antemão, que as pesquisas não seria passível de publicação ampla. (AVALIADOR MSF4)

> Mas isso é o que pode acontecer em uma avaliação de impacto. Nem sempre você vai achar um impacto positivo. Mas isso é uma avaliação de impacto, né? Você pode alcançar um resultado positivo ou não. O resultado que não é positivo também é um resultado. (AVALIADOR MSF1, grifo da autora da dissertação)

> Então eu acho que foi adequada sim a avaliação de impacto. Embora, muitas vezes, a gente fique procurando um impacto positivo e nem sempre a gente encontra. (AVALIADOR MSF1)

Na perspectiva dos avaliadores da FJP, as avaliações trariam novas possibilidades para órgãos como a FJP e a EMATER, de acordo com a fala do avaliador MSF2. Para o centro de pesquisas, as avaliações ocasionariam a introdução de uma nova seara de atuação para desenvolver e aprimorar políticas públicas; e para a empresa de assistência rural, um processo de conscientização da importância de entregar resultados, e não produtos, para a sociedade. Na percepção do avaliador MSF2, no início do processo de avaliação, as promessas do trabalho seriam mais abstratas do que concretas, mais possibilidades do que fatos.

> Era ainda muito mais um potencial do que algo concreto, era de você trazer uma perspectiva nova de pesquisa para a avaliação para a Fundação João Pinheiro, que era avaliação de políticas públicas. E não só para a Fundação João Pinheiro e também para esses órgãos, porque, na medida em que a EMATER, que outros estavam envolvidos, acho que havia um processo de conscientização de que não basta você fazer entregas, que essas entregas têm que ter resultados efetivos para a sociedade e que não é tão simples você conseguir captar isso. (AVALIADOR MSF4, grifo da autora da dissertação)

Para o avaliador MSF2, apresar de não haver resistência por parte da EMATER, no início do processo existiu uma apreensão por parte da equipe do projeto. O receio dos usuários da avaliação, naquele momento, era de que a avaliação não conseguisse captar o ganho social do projeto e mudanças nas condições de vida do beneficiário. No entanto, de acordo com o avaliador v, esse *medo inicial* não atrapalhou o desenvolvimento da avaliação.

> Agora, medir esse ganho, medir é o impacto disso aí lá na vida social do agricultor, lá na casa dele, na alimentação dele, na melhoria das condições de vida, na melhoria habitacional, aí é mais difícil, entendeu? (EQUIPE DO PROJETO MSF1)

> Não teve uma apreensão... bom, no começozinho teve esses tipos de apreensão. Tipo, "olha, vocês tem que tomar cuidado porque, né? A gente sabe que a coisa dá certo se for feita do jeito certo". Eles tinham essa impressão geral e uma certa apreensão de que a avaliação num fosse levar essas coisas em conta, mas depois que passado esse medo inicial num teve muito problema não. (AVALIADOR MSF2)

Observa-se, assim, que o contexto em que a avaliação do projeto MSF se inseriu foi um contexto de receptividade por parte dos usuários da mesma, com a expectativa de que o estudo apontasse pontos de melhoria e os resultados promissores do projeto. O processo de implementação e execução da avaliação, no entanto, não foi tranquilo. Verificam-se

dificuldades na seleção das abordagens avaliatórias e desenvolvimento da pesquisa como apontado na seção a seguir.

9.2 Implementação do processo de avaliação e envolvimentos dos atores

Na tentativa de remontar o processo de implementação da avaliação do MSF, o gestor do projeto divide facilmente essa etapa em três partes: primeiro, a realização da análise documental pelos avaliadores das informações disponíveis sobre o projeto para entendimento do seu contexto de atuação; em um segundo momento, a execução das entrevistas com os funcionários da EMATER que trabalhavam diretamente no projeto e, por fim, a efetivação da pesquisa de campo para a coleta de dados. Embora essa divisão simplificada possa demonstrar, à primeira vista, um processo de avaliação sossegado, sem muitas dificuldades aparentes, oculta, contudo, os vários problemas de desenho da avaliação do MSF, dificuldades em relação aos dados e ao envolvimento dos atores envolvidos. A percepção ingênua do entrevistado pode ser identificada no trecho abaixo.

> A pesquisa foi dividida em três partes. Uma parte de análise documental. Outra parte de entrevistas com os gestores daqui, com o diretor do programa, e depois a execução de campo. Feito isso, eles retornaram e discutiram conosco os resultados preliminares da pesquisa. Após a apresentação, voltaram, vamos dizer assim, para o laboratório para escrever o relatório e apresentar para o governo que era o encomendante da pesquisa. Na verdade, o encomendante é era a SEPLAG. (GERENTE DO PROJETO MSF)

A implementação da avaliação do MSF não foi tão tranquila quanto relatado pelo gerente do projeto. Ao logo da avaliação, o projeto de avaliação inicial foi perdendo o seu desenho original e se transformando em outra pesquisa, devido a fatores como ausência de informações e dados sobre o MSF e problemas de diálogo entre os atores. De acordo com o avaliador MSF3, o desenho da avaliação foi construído ao longo do tempo, sendo reformatado quando necessário, e não de uma única vez, com um projeto de avaliação estável.

> Foi um processo que foi sendo construído ao longo do tempo, ele não foi pensado, ele não foi formatado "Ah, vai ser dessa forma. É assim que vai ser". Ele foi sendo acontecendo, ele foi sendo adaptado às condições que a gente tinha, seja de recurso, seja dos próprios dados que a EMATER tinha para nos fornecer. (AVALIADOR MSF3)

O primeiro contato para a execução da avaliação foi realizado pela SEPLAG com a presença do consultor contratado pela FJP. Na visão do entrevistado, o cliente principal da avaliação do

MSF seria a SEPLAG, responsável pela contratação da avaliação. Observa-se que o gestor do projeto não cita a presença do EpR em nenhum momento e tem dificuldades em diferenciar o EpR da SEPLAG, conceituando as duas estruturas como uma única. É possível verificar também, a partir da fala do gerente do projeto MSF, uma delimitação de trabalho muito clara entre o trabalho do consultor, em conjunto com a SEPLAG, e as atividades da equipe de avaliadores da FJP, como se o consultor e a FJP nem fizessem parte de um mesmo time de avaliação. O gerente limita o trabalho do consultor e da SEPLAG até o desenho da avaliação e, a partir daí, conta com a presença dos avaliadores da FJP. Nota-se que, na visão do gerente, o trabalho da FJP começa quando o trabalho do consultor e da SEPLAG termina.

> Nós tivemos algumas reuniões preparatórias com a equipe da Fundação João Pinheiro, eles tinham um consultor externo da Fundação Getúlio Vargas, nós também tivemos algumas reuniões com ele. Primeiro, veio a SEPLAG e o consultor para montar o arcabouço, delimitar o escopo, definir o que poderia ser avaliado, porque o programa é muito grande e eles decidiram que não iriam fazer a avaliação dele no total. Iriam avaliar no geral e focar em algumas ações. Então, nós trabalhamos com a SEPLAG, com o consultor para definir o que seria avaliado. Depois, o consultor entendeu essa fase como encerrada e passamos a lidar diretamente com a Fundação João Pinheiro. Aí, com a Fundação João Pinheiro, nós montamos um plano de ação, identificamos o gestor de cada ação aqui e depois ir a campo fazer um trabalho lá com o pessoal da execução da avaliação e com o público alvo do programa. (GERENTE DO PROJETO MSF)

Além disso, na visão do gestor, a participação do consultor contratado da Fundação Getúlio Vargas não foi essencial ao desenvolvimento do projeto de avaliação do MSF. Como o contato da EMATER foi mais estreito com a equipe de avaliadores da FJP, do ponto de vista do usuário, o consultor contratado não teve tanta importância no processo.

> Do meu ponto de vista, do ponto de vista de usuário da avaliação, eu não senti muito a presença da Fundação Getúlio Vargas, da consultoria técnica. Eu tive mais contato com o pessoal da Fundação mesmo. No decorrer da pesquisa mesmo eu não senti, assim, tipo assim, vamos dizer assim, que não fizeram falta não. No decorrer da pesquisa eu não vi muito o dedo dessa consultoria. (GERENTE DO PROJETO MSF)

Tendo em vista o relato do gerente do projeto, é possível observar que a expectativa original de criar capacidade de avaliação com o trabalho dos consultores – como discutido no capítulo seis, estimulando o aprendizado entre a equipe de avaliadores da FJP, não pôde ser identificada na avaliação do MSF. Outro ente que não teve presença assumida na avaliação foi a SEAPA, secretaria de estado à qual a EMATER se vincula. O processo de avaliação não contou com a participação da SEAPA, como relatado pelo gerente do projeto: *na verdade, a*

SEAPA participa assim, como conhecedora do programa, mas não tem uma ação muito específica.

Em oposição à distancia de atores como o consultor contratado e a própria SEAPA, o envolvimento dos avaliadores da FJP com os técnicos da EMATER foi frequente segundo os relatos das entrevistas. Todos os avaliadores expuseram a participação e envolvimento entre as equipes. Para Johnson *et al.* (2009), o envolvimento dos *stakeholders* com a avaliação deve ser uma preocupação recorrente em estudos avaliatórios e impactam positivamente o uso das avaliações. Alguns relatos exemplificam a cooperação entre avaliadores e avaliados no processo do MSF.:

> Estavam, estavam abertos. O que precisou de informação foi prestado. E olha, a gente fez diversas reuniões longas, porque tem todo um linguajar e tem toda uma especificidade que é da área rural, de agronomia, de veterinária, estão para nos ajudar nisso eles nunca tiveram problema. Quando fomos chamados, foram à Fundação João Pinheiro, quando precisou a gente foi lá. Por exemplo, questionário, que é uma parte bem complicada, não sei como isso evoluiu, mas a gente fez um questionário bem grande e para você acertar esse questionário é difícil, porque nós não temos esse linguajar do homem do campo. Então, a gente também fez esse trabalho com eles de sentar e passar cada pergunta e eles tentarem: "Tá, mas o que vocês querem aqui?" ou "Mas vocês não podem perguntar assim, vocês tem de perguntar assado...". (AVALIADOR MSF4)

> A EMATER deu muito apoio para a gente. Ela foi uma grande parceira. Tudo que a gente precisou a gente pode contar com ela. Ela ajudou muito na elaboração do questionário, né? A gente elaborou um questionário, levou, discutiu com eles várias vezes. Depois, no campo mesmo, quando a gente foi a campo eles nos ajudam, agente tinha apoio no escritório da EMATER. Alguns dados que eles tinham lá dentro eles disponibilizaram para gente. Dentro do possível eles nos ajudam muito. Dentro do que eles podiam, né? Porque tinha coisas que a gente precisava que eles não tinham como nos ajudar. (AVALIADOR MSF1)

Na versão do gestor do projeto, a percepção dos avaliadores também não estava errada. Segundo o relato abaixo, havia um interesse por parte da equipe do MSF em ter uma avaliação que apontasse seus problemas e oportunidade de melhoria e, devido a isso, não houve resistência por parte da EMATER em repassar os documentos e dados à equipe de avaliação da FJP em nenhum momento da avaliação. Essa abertura, comentada, tanto pelos avaliadores, quanto pelo próprio gerente do projeto, remete ao fato de não enxergar a avaliação como algo fiscalizatório na percepção do gestor.

> Em momento algum nós nos sentimos investigados. Isso foi uma coisa que até dificulta, né? Se você se sente investigado, vamos dizer assim, vou esconder uma coisinha debaixo do tapete. Não tinha nada a esconder e tínhamos todo o interesse

> de que as eventuais mazelas do programa fossem detectadas para poder corrigir. (GERENTE DO PROJETO MSF)

A mesma receptividade identificada entre os avaliadores da FJP e os funcionários da EMATER não pode ser observada, contudo, na relação entre os avaliadores e os técnicos e gestores do EpR. Na ótica dos avaliadores, o EpR (às vezes entendido como SEPLAG) tinha o papel de *intromissão mesmo* na visão dos avaliadores MSF2 e MSF4. Na visão desses avaliadores o EpR (que também é chamado de SEPLAG por um dos avaliadores) estava preocupado em controlar as decisões do processo de avaliação e, posteriormente, o andamento da mesma – cronograma, produtos e entregas. Já, na visão do avaliador MSF3, apesar de o EpR acompanhar o projeto, as definições importantes sempre partiram da equipe da FJP.

> Acho também que a questão da SEPLAG, do escritório lá demandando isso dava mais um componente externo, e, eu, pelo menos (estou falando em meu nome) não me sentia tão livre, com tanta autonomia para conduzir meu trabalho de avaliador e algumas vezes a própria Fundação teve que se posicionar. Papel de intromissão mesmo! Se eu tenho que remarcar com EMATER eu tenho que avisar para a SEPLAG? Por que eu tenho que avisar para a SEPLAG? Eu tenho que ter liberdade para conduzir o meu trabalho. Então, esse tipo de coisa. Eu acho que a SEPLAG queria controlar muito o processo de avaliação. (AVALIADOR MSF4)

> Mudava um pouco as pessoas que acompanhavam o projeto, mas sempre teve alguém na época do EpR que acompanhava o projeto. Então, toda vez que havia uma definição do que pretendia fazer, a gente apresentava tanto para EMATER e sempre na presença de alguém do EpR. Mas a definição, o que seria feito sempre partiu da própria Fundação, da equipe, os outros participavam muito mais no sentido de contribuir do que indicar caminhos mesmo. Assim, quem definia "vamos fazer determinado tipo de avaliação" sempre partiu muito mais da equipe. (AVALIADOR MSF3)

O processo de detalhamento e definição das abordagens metodológicas para o desenvolvimento da avaliação também não foi trivial de acordo com os relatos. Observa-se que não existiu, no caso da avaliação do MSF, um projeto de avaliação fixo, desenhado pelo consultor contratado. A ideia inicial da avaliação, de acordo com os relatos, seria fazer uma avaliação do impacto geral do projeto, de todas as diferentes ações do projeto que, de acordo com os avaliadores, seriam quatorze ações variadas. No entanto, muitas mudanças foram realizadas até se chegar à seleção final das três frentes do projeto. À medida que os avaliadores se aproximavam mais do projeto, esbarravam em limitações metodológicas e de informações. O trabalho não foi nada trivial de acordo com os relatos.

> Até chegar nessas três frentes não foi fácil não, sabe. Até a gente conseguir fechar teve muita reunião para discutir. Até mesmo para decidir qual ação seria avaliada, porque eram 14 ações, né? não tinha como fazer uma amostragem para todas essas

> ações e fazer uma avaliação de impacto para todas essas ações então a gente teve que eleger quais ações seriam avaliadas e de que forma. (AVALIADOR MSF1)

> Inicialmente tinha espectro enorme, se pensava em avaliar, fazer entrevista em profundidade, e que à medida que foi sendo lapidada essa avaliação viu-se que muita coisa saiu muito fora ou então, às vezes, tinha até dificuldade de implementar realmente.(AVALIADOR MSF4)

> Quando eu entrei tinha uma proposta de fazer avaliação de todas as ações de várias formas. Eu não sei te detalhar direito, mas eu lembro que era um escopo muito maior do que no final das contas a gente acabou trabalhando. O projeto, ele foi mudando à medida que a gente ia se deparando com as limitações principalmente do banco de dados e também de limitação de recurso. (AVALIADOR MSF3)

Depois de muita discussão entre os avaliadores da FJP para realizar a avaliação de impacto do projeto, foram eleitas três ações do projeto: lavouras, apicultura e tanques de resfriamento de leite. Em 2009, foi realizada a avaliação dos critérios associados à distribuição das ações de lavouras temporárias, apicultura e tanque de resfriamento de leite. No primeiro semestre de 2010, foi feito estudo abordando a viabilidade econômica financeira das ações de tanque de resfriamento de leite e apicultura. A ação lavoura, que contempla 28% das famílias beneficiadas e para qual se destina o maior percentual dos recursos financeiros do MSF (44%), constituiu objeto da terceira etapa do projeto de avaliação do projeto. Para a avaliação de impacto da ação de lavoura, foram considerados dois indicadores: a renda *per capita* domiciliar e a segurança alimentar domiciliar, e os resultados apresentados neste relatório dizem respeito à avaliação de impacto nesses indicadores (FUNDAÇÃO JOÃO PINHEIRO, 2010b).

Observa-se, também, que o processo da avaliação do MSF foi dificultado pela ausência de dados e pela falta de clareza das informações do projeto, de acordo com o depoimento dos avaliadores. Lobo (2001) ressalta, desde o desenho da política pública, a importância da informação. A estruturação de sistemas disponíveis, de acessibilidade garantida e atendendo às peculiaridades dos programas sociais, é fundamental para o mínimo de factibilidade das avaliações acrescenta Lobo (2001). Assim, sistemas de dados, bem estruturados, desde o início do projeto, possibilitam estratégias de avaliação mais condizentes com seu objetivo. Contudo, no quesito dados, a avaliação do MSF se defrontou com inúmeros problemas detectados na fala dos avaliadores.

> Os problemas que a gente teve, por exemplo, com o cadastro deles é porque o cadastro não era pensado para avaliação de impacto. Inclusive, uma das sugestões, considerações que foram colocadas depois era isso, uma preocupação que eles

> deveriam ter com a qualidade do cadastro para que eles próprios depois pudessem utilizar esse cadastro. Porque eles realmente eram muito deficientes. (AVALIADOR MSF3, grifo da autora da dissertação)

> Quando o banco de dados da EMATER chegou até aqui com o cadastro dos beneficiários era um banco de dado muito deficiente, porque não era objetivo; ele não é feito para nenhum objetivo de avaliação, é simplesmente um cadastro que eles têm de beneficiários. Então muito do que se tinha pensado, na hora que você via o cadastro não era possível ser feito. Você fica fazendo adaptações para que você consiga fazer a avaliação. (AVALIADOR MSF3)

> A gente tinha que trabalhar com os dados que eles tinham, né? E as vezes seria melhor se a gente tivesse outros tipos de informação também. Mas não tinha jeito, a gente tinha que trabalhar com que estava disponível. Tem o ideal e o possível, né?! (AVALIADOR MSF1)

Ainda, na percepção dos avaliadores, os problemas de dados estariam também relacionados a problemas metodológicos das avaliações de impacto, quais sejam, a dificuldade de planejar avaliações de impacto após a implementação dos projetos e a predominância de aspectos quantitativos das avaliações. Em relação ao primeiro problema, os avaliadores ressaltaram que, caso a avaliação de um programa governamental fosse planejada antes da intervenção do mesmo ser iniciada, não seria necessário resgatar todo o processo de planejamento, uma vez que as populações beneficiadas e os próprios dados gerados seriam prerrequisitos para uma avaliação confiável do programa. No entanto, a cultura imediatista do setor público brasileiro, somado a interferências políticas existentes, impede esse tipo de prática, como comentam os avaliadores nos trechos abaixo.

> A fraqueza é essa: pensar em uma avaliação quando o programa já caminhou muito. Eu acho que uma avaliação de impacto, eu acho que ela pode trazer melhores resultados se ela for pensada antes do programa começar. Mas isso é muito difícil de acontecer, né? Geralmente o que acontece é você pensar em avaliar um programa depois que ele já está implementado dois, três anos. E aí você fica com uma dificuldade de definir um grupo controle. (AVALIADOR MSF1)

> Uma coisa é você fazer, já pensar na avaliação quando o projeto está sendo desenhado e outra coisa é você pensar fazer uma avaliação quando o projeto já está a anos no caminho. Por quê? Porque daí tem um trabalho de reconstrução do que foi o planejamento desse projeto ou até do que não foi o planejamento desse projeto. Então é quase como você tem que resgatar todo o processo de planejamento junto com quem participou disso para você conseguir desenhar a política pública. E tem também a questão da linha de base, que, no caso, para nós não estava dado, então a metodologia seria exatamente você selecionar o grupo beneficiário e o grupo similar não beneficiário. (AVALIADOR MSF4, grifo da autora da dissertação)

Em relação ao aspecto, à complexidade inerente às avaliações de impacto, o avaliador MSF3 protesta contra a predominância do modelo quantitativo nas avaliações. Fato já comentado por

teóricos do campo, como Guba e Lincoln (1989), a constituição do campo de avaliações de programas caracterizou-se por pressupostos típicos do rigor positivista, cuja influência se mostra ainda presente acompanhando o histórico da administração pública do País, como ressalta Lobo (2001). Nesse sentido, o avaliador MSF3 argumenta que *não necessariamente a avaliação de impacto é a resposta para todas as avaliações*, sendo a combinação de métodos qualitativos e quantitativos uma alternativa interessante para explorar os resultados e os processos de um projeto.

> Não necessariamente a avaliação de impacto é a resposta para todas as avaliações. Algumas pessoas questionaram a avaliação de impacto, eles achavam que era algo muito mais matemático, muito frio e que talvez não conseguisse captar tudo. Eles achavam que talvez uma metodologia mais qualitativa conseguisse perceber mais outras nuances do projeto. Eu, particularmente, gosto muito da avaliação qualitativa. Eu acho que tem hora que você ir e conversar com a pessoa e tentar perceber como que ela vê o projeto e quais são as sugestões que elas dão, eu acho que isso pode ser as vezes muito mais interessante do que simplesmente fazer uma avaliação do antes e do depois, porque eu acho que durante é importante também. Eu acho que o ideal é você fazer uma "quali" antes e depois uma "quanti" ou vice-versa, porque eu considero que elas são complementares. (AVALIADOR MSF3, grifo da autora da dissertação)

De acordo com o entrevistado do EpR (Coordenação EpR1), apesar de todas as avaliações coordenadas pela instituição sofrerem mudanças dentro do intervalo de tempo entre o desenho inicial, preparado pelos consultores contratados (relatório de viabilidade), e o projeto de execução da avaliação, elaborado pela equipe de avaliadores FJP de cada projeto, tais mudanças foram mais intensas no caso da avaliação do MSF. Na visão do entrevistado, no MSF, as mudanças, que foram mais profundas e prejudicaram o andamento da avaliação, foram causadas devido a problemas de diálogo entre o consultor contratado e a equipe de avaliadores da FJP.

> Entre o relatório de viabilidade e o, digamos, relatório de metodologia você tinha mudança, porque você tinha uma visão do que poderia ser feito e aí você amadurecia até a visão do que pode ser feito. Geralmente a mudança era pequena, ou seja, era mudança muito mais de ajuste de modelo quando você vê os dados, a qualidade dos dados, coisa do tipo no nível diretório maior. Mas aqui a distância era pequena, exceto o Minas Sem Fome, porque você tinha uma metodologia e aí você viu que, na verdade tinha um problema nessa entrega porque não se discutiu o suficiente entre consultor e equipe, e aí entregou alguma coisa ou o consultor aceitou alguma coisa e a equipe... E no final a equipe não achou aquilo legitimo e fez uma quinta coisa, uma quarta coisa que tinha um descolamento. Então, ele teve um problema de defensão do que era o escopo do Minas Sem Fome. Definiu, mas a equipe não achou legitimo e mudou depois. Que é o problema de entendimento do que o consultor faz. Ou seja, a equipe que se acha mais esperta e acha o consultor desnecessário, ela faz esse tipo de coisa. Se aceita um produto que o consultor faz e fala: "Eu vou fazer melhor. Eu vou fazer do meu jeito". A pergunta é: então para que contratou um consultor? Mais que tem a ver com esse processo de rebeldia. E um perfil do

> consultor também que não ajudou tanto, que não era alegador, que era muito atarefado, pegava muita coisa. Tem essas coisas todas. A ideia era nobre e a implementação não foi tão fez assim por motivos da resistência interna, um consultor não foi tão bom do ponto de vista dessa interação com a equipe. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

Na analogia de Guba e Lincoln (1989), os avaliadores são como orquestradores, maestros, de um processo de negociação que objetiva alcançar o consenso entre uma construção viável da avaliação e a melhor informação gerada possível. No entanto, para conseguir tal objetivo, os avaliadores devem agir como parceiros dos diversos *stakeholders* envolvidos em uma avaliação (GUBA; LINCOLN,1989). Observa-se, no entanto, que o processo de comunicação entre a própria equipe de avaliadores e o consultor contratado não conseguiu chegar a um consenso. Na visão do entrevistado EpR1, a equipe de avaliadores experimentou um processo de rebeldia, de resistência, quanto às ideias do consultor contratado e quem perdeu foi a avaliação do MSF.

> Aí perdeu o foco. De novo, tudo por falta de uma liderança, digamos, de padrão. Ou seja, esse é o grande risco, "Ah, então eu vou avaliar o que eu sei avaliar. Eu vou avaliar um trenzinho qualquer". Isso é o que o governo quer? Isso é o que o governo precisa? De novo, pode até precisar e pode ser até importante saber se instalar um tanque de leite dá retorno ou não, mas não era o *script* original e a coisa foi desgringolar para isso e foi se deixando desgringolar para isso por falta dessa liderança. (COORDENAÇÃO EPR1)

Esses depoimentos salientam a importância do avaliador, além dos estudos já citados, outros trabalham devotam atenção para o avaliador. O trabalho de Johnson *et al.* (2009) aponta que os profissionais da avaliação desempenham um importante papel na condução de sua utilização porque os avaliadores devem, dentre variadas funções, articular interesses e valores diversos inseridos em uma avaliação, respeitar os padrões técnicos adequados aos métodos por eles selecionados, certificar que existe uma ampla compreensão dos envolvidos quanto ao contexto da avaliação, garantir integridade do processo de avaliação por exemplo. De acordo com os autores, a característica do avaliador, sua competência ou liderança têm uma natureza influenciadora forte na condução da avaliação e, consequentemente, no momento de utilizar os resultados de uma avaliação (JOHNSON *et al.*, 2009). No caso do MSF, ainda que a avaliação fizesse sentido entre os avaliadores, ficou inconsistente do ponto de vista da coordenação do processo, EpR que, por sua vez, considerou a avaliação ilegítima frente à demanda inicial.

É possível notar, assim, problemas significativos no processo de implementação da avaliação do MFS que podem impactar negativamente as avaliações. O primeiro deles estaria relacionado à dificuldade de delimitação do estudo de impacto do projeto que repercutiu em uma estratégia diferenciada da inicialmente negociada com a coordenação do projeto de avaliações no EpR. Enumeram-se também os problemas de natureza metodológica como, por exemplo, as dificuldades relacionadas à seleção do grupo-controle e grupo-tratamento e de resgate dos dados beneficiários para a mensuração do impacto. Por fim, torna-se importante ressaltar a falha de legitimidade recíproca entre EpR e grupo de avaliadores da FJP, sendo o gestor do projeto alheio à ausência de consenso entre eles. As implicações desses problemas aqui ressaltados são então repercutidas nos processos de processamento de informações da avaliação, por parte de seus usuários, como discutido na seção seguinte.

9.3 Resultados e processamento de informações

O processo de comunicação dos resultados e o processamento das informações pelos usuários de uma avaliação são considerados fatores essenciais para um bom aproveitamento das avaliações. Leviton e Hughes (1981) tratam do processamento de informações como o reconhecimento dos resultados de uma avaliação e a sua tradução em sínteses, tópicos de recomendações, entre outros produtos, para melhor compreensão de ações implementáveis advindas da avaliação. De acordo com os autores, a maneira como os resultados das avaliações são apresentados aos usuários afeta diretamente a sua compreensão e, assim, a extensão de seu uso (LEVITON; HUGHES, 1981). A mesma opinião dos autores é refletida na fala do avaliador do MSF, embora ele reconheça que se trata de um trabalho *cansativo* para o avaliador.

> O papel da comunicação do avaliador. E esse papel é importante também, embora cansativo pro avaliador, porque o avaliador já fez um trabalho grande ali, né? Não é fácil fazer uma avaliação e tudo mais, e aí você tem esse trabalho de resumo, de comunicação, de passar aquilo da melhor maneira. Mas é um trabalho importante. (AVALIADOR MSF2)

A dinâmica de entrega dos resultados, de acordo com o avaliador MSF3, foi feita a cada novo produto concluído na avaliação e, não, de uma única vez, como aconteceu em outras avaliações do conjunto.

> Nós fizemos uma apresentação final a cada produto, quando ele concluía, a gente fazia uma apresentação. E estas apresentações sempre com o pessoal da EMATER, com o pessoal do EpR e com a equipe daqui. Sempre tivemos apresentações formais e, depois, a gente entregava o produto para eles avaliarem. Inclusive, a gente apresentava, deixava o material com eles, depois eles marcavam uma nova reunião. (AVALIADOR MSF3)

Desde o início da avaliação até a entrega dos relatórios finais, os avaliadores estimaram dois anos e meio de trabalho. O atraso na entrega dos relatórios foi de dois meses. De acordo com os avaliadores, o relatório estava previsto para outubro e foi concluído e entregue em dezembro do mesmo ano, 2010. Para os usuários, esse atraso não trouxe problemas na utilização dos resultados. Para os entrevistados da EMATER, a qualquer tempo, os resultados da avaliação seriam tempestivos.

> Pra mim não existe hora certa. A hora certa é aquela que você tem acesso à informação, trabalha essa informação e coloca ela justamente dentro do nosso trabalho pra melhoria. (EQUIPE DO PROJETO MSF1)

> Eu devo confessar para você que eu não me lembro se demorou não. Mas é o tipo de avaliação que ela também não tem data marcada, então, eu considero que foi positivo. Não me lembro de ter atrasado no cronograma ou não, mas chegou em tempo hábil, e a qualquer momento ela iria ser importante. Se fosse feita hoje, teria validade hoje.... Aliás, uma coisa que eu imagino é que deveria ser repetida. (GERENTE DO PROJETO MSF)

Os principais resultados da avaliação de impacto do programa MSF, no que tangue à ação lavouras, apontaram tantos pontos positivos, bem como algumas questões que necessitam de melhoria. Um dos apontamentos positivos do projeto diz respeito à eficácia da implementação da ação de lavoura, a partir da conclusão que a distribuição de sementes aos beneficiários do programa MSF produz um impacto na renda domiciliar *per capita* dessas famílias de, aproximadamente, R$25,00[11]. Observaram-se, no entanto, pontos negativos quanto à cobrança do associativismo para participação no projeto, já que o beneficiário tem que participar de alguma associação para receber o benefício; ao conhecimento do projeto pelos próprios beneficiários do MSF e, por fim, ao impacto negativo na segurança alimentar das famílias que participam do projeto, apesar de a distribuição de sementes aumentar a probabilidade de as famílias se preocuparem com a segurança alimentar, não chega a impactar, de forma quantificável, a segurança alimentar em si (FUNDAÇÃO JOÃO PINHEIRO, 2010b). Os

[11] Resultado controlado por características do chefe, composição do domicílio, condições de moradia, participação em associação, recebimento de outros benefícios sociais e conhecimento do programa.

avaliadores comentaram os resultados da avaliação de impacto da ação lavouras, destacando a falta de efetividade do projeto quanto ao objetivo de promover a segurança alimentar.

Os resultados das avaliações econômicas das ações de distribuição de tanques de leite e de apicultura foram melhores que da ação lavoura. Os resultados apresentados sugeriram a viabilidade econômica das duas vertentes do programa analisadas: do incentivo à apicultura e do programa dos tanques de expansão para refrigeração do leite (FUNDAÇÃO JOÃO PINHERO, 2010d). Contudo, mesmo apresentando avaliação positiva, os relatórios dessas duas ações específicas contaram com variadas recomendações sobre aspectos de gestão das ações do projeto.

Observa-se, dessa maneira, que a avaliação de impacto do projeto MSF não apresentou resultados positivos do projeto, como era esperado pelos entrevistados da EMATER, mas, sim, indicativos de melhoria. Para Cousins e Leithwood (1986), quando os resultados da pesquisa apontam questões diferentes das crenças e expectativas dos usuários, isso pode trazer complicações ao uso e à receptividade. Para os autores, descobertas esperadas estariam mais relacionadas à aceitação e à utilização por partes das equipes dos programas, enquanto resultados surpreendentes das avaliações estariam fadados à resistência e ao não uso. No entanto, de acordo com o relato de avaliadores, independentemente dos resultados dos três estudos avaliatórios, houve abertura por parte da EMATER.

> E quando a gente levou os resultados finais e apresentamos para eles, até para o próprio presidente da EMATER, você percebia isso, que aquilo que a gente estava apresentando para eles era muito interessante porque era uma forma de melhorar o trabalho deles. (AVALIADOR MSF3)

> A gente fez uma boa comunicação pra eles dos resultados e aí que eles se mostraram extremamente abertos. Foram duas reuniões de comunicação. Uma com o presidente e a outra antes do presidente, tá? E eles gostaram, viram que o trabalho era técnico, era sério, era pra avaliar o Programa mesmo, né? (AVALIADOR MSF2)

Esse discurso é corroborado pelo relato do gestor do projeto que confirma a percepção dos avaliadores na medida em que se diz satisfeito com a avaliação elaborada pela FJP. De acordo com o gerente do MSF, a equipe do projeto estava preparada para receber mais pontos negativos do projeto que pontos positivos, pois os *defeitos* do projeto são passíveis de correção e são os pontos que merecem a atenção da equipe. A satisfação do gerente com a avaliação é demonstrada em pontos diferentes da entrevista.

> Foi até bem recebido, até porque tinha esse espírito de receber. Pontos falhos certamente têm, todo processo tem. Mas a gente se preparou para receber. O estado da arte é uma coisa e a vida real é outra. Então o adequado numa pesquisa desse tipo é você estar preparado mais para ter defeito do que para ter qualidade. Qualidade é decorrente, o defeito você consegue trabalhar em cima para corrigir. Se eu fosse encomendar uma avaliação hoje, eu acho que não mudaria muito não. Eu acho que ela conseguiu chegar no objetivo proposto. Porque eu acho que o fundamental com uma avaliação desse tipo é a isenção. E outro é o próprio procedimento de campo. Hora nenhuma nós palpitamos nisso aí, de ter algum tipo de ação que pudesse inibir a intenção da pesquisa. E uma pesquisa que não é isenta não é pesquisa, não adianta. (GERENTE DO PROJETO MSF, grifo da autora da dissertação)

De acordo com o avaliador MSF2, a comunicação dos resultados das avaliações foi positiva. Houve um interesse por parte do usuário em conhecer os bancos de dados, em confeccionar resumos entre outras ações. No entanto, entre a receptividade com os resultados da avaliação e a implementação da mudança em si é um longo caminho. Apenas receber bem a avaliação não quer dizer que haverá um uso instrumental, como comentado na fala do avaliador e como tema da seção seguinte.

> E a comunicação foi positiva, tanto é que eles pediram pra passar o banco de dados, né? Pediram pra ler, é eles leram tudo, ficaram a par, gostaram bastante da avaliação foi bem recebida. Ai entre a avaliação ser bem recebida e ela ser usada pra alterar o dia a dia do Programa, pra alterar as atividades e tudo mais é que eu acho que isso se perdeu. Sei que eles gostaram, mas não sei como eles aplicaram, entendeu? Então não tem o seguimento disso. (AVALIADOR MSF2, grifo da autora da dissertação)

9.4 Utilização da avaliação do ponto de vista dos *stakeholders*

De acordo com um dos entrevistados da avaliação do projeto Minas Sem Fome, a grande dificuldade das avaliações ocorre após a comunicação dos resultados: *a parte depois da comunicação do resultado e o uso que é o grande problema dessa avaliação*. São variadas as percepções da utilidade das avaliações para o MSF dependendo do tipo do *steakeholder* e do contato que ele teve com a avaliação. Assim, observam-se visões diferenciadas da utilização para os avaliadores da FJP; para os entrevistados do EpR, que estavam na coordenação do projeto e para o gerente do projeto, como seu usuário principal.

Após a conclusão da avaliação, os próprios avaliadores não souberam relatar qual foi o uso da avaliação do MSF. De acordo com o avaliador MSF3, não é de conhecimento se as recomendações colocadas nos relatórios repercutiram em mudanças no projeto. Os avaliadores MSF1 e MSF2 arriscam o palpite de que o uso foi precário.

> Realmente eu não sei se as recomendações, os resultados de cada uma das avaliações, se isso refletiu internamente em algum tipo de mudança: se melhoram a qualidade do cadastro; se passaram a ter um cuidado maior com esse cadastro. Isso eu não sei porque realmente a gente não teve esse *feedback*. (AVALIADOR MSF3, grifo da autora da dissertação)

> Eles não tiveram muita chance porque o programa estava, assim, com grandes chances de não continuar. E eles estavam mais preocupados em conseguir manter o programa porque em 2011 eles tiveram uma certa dificuldade neste sentido. Então, eles não tiveram muita chance de trabalhar os resultados da avaliação. Porque eles estavam preocupados em conseguir que o programa continuasse. (AVALIADOR MSF1)

> A gente não sabe como se deu o uso. Entregamos, aí comunicamos e agora o resultado é de vocês e... ficou por isso, entendeu? Tipo, havia uma intenção de fazerem novas rodadas de avaliação, mas, como começa a coisa, a equipe se dissipou, a coisa foi mudando e lá no EMATER também eles tiveram várias dificuldades, num teve acompanhamento posterior e nem finalização de acompanhamento posterior, né? O ideal seria nós aqui mesmo também fazer um projeto de avaliação, mas a coisa perdeu menos, como o Banco Mundial se retirou um pouco disso, né? A coisa perdeu um pouco o ânimo. Então eu acho que o uso da avaliação é um ponto que foi precário. (AVALIADOR MSF2, grifo da autora da dissertação)

Os motivos para a não utilização, de acordo com o avaliador MSF1, têm a ver com razões de contexto político e do momento em que se encontrava o governo. De acordo com o relato, entre 2010 e 2011, ocorreu a troca do governo mineiro e, apesar de o partido continuar o mesmo, algumas mudanças de ordem orçamentária impactaram os projetos estruturadores, dentre eles, o MSF. Já, na visão do avaliador MSF2, a equipe de avaliadores da FJP se dissipou após a conclusão da avaliação, devido à rotatividade presente na instituição, em particular nos grupos que trabalhavam no Centro de Estudos de Políticas Públicas, setor que estava cuidando das avaliações naquele período. Outro desmotivador, para o avaliador MSF2, foi a saída dos técnicos do Banco Mundial, cuja presença foi essencial no momento de desenho das avaliações. A impressão que ficou, de acordo com o avaliador MSF1, foi de que o tipo de avaliação de impacto, após o reconhecimento da ausência de impacto dos projetos avaliados, *saiu de moda*; o que antes da realização das avaliações era considerado tendência entre as avaliações, no momento de sua conclusão já não era mais prioridade, como se sugere no comentário abaixo.

> O que eu tenho dúvida é se esse tipo de avaliação de impacto tem sido tão bem recebida como já foi. Anteriormente a avaliação de impacto estava meio que na moda e agora, com os resultados que vieram dessas avaliações de impacto, que não foram tão positivos quanto o esperado. Não que não foram positivos. Essa moda meio que saiu de moda. Tem se buscado outros tipos de avaliação, até mesmo

> qualitativas, na busca de um efeito dos programas que, muitas vezes, não é conseguido com avaliação de impacto. Mas isso não é uma coisa específica do MSF, é uma coisa de contexto mesmo. É geral. (AVALIADOR MSF1, grifo da autora da dissertação)

Na visão dos gestores do EpR, a avaliação do MSF foi *mais protocolar*, uma vez que a instituição não teria capacidade e nem cultura para absorver a avaliação. Nas palavras do entrevistado, não estaria *introjetado no DNA* da EMATER esse tipo de trabalho. Assim, a EMATER teria aceitado a avaliação como uma obrigação e, não, como um processo de melhoria contínuo estimulado pela sua utilização.

> No Minas Sem Fome acho que foi mais protocolar, aí a EMATER já não tinha capacidade de tocar um processo, a equipe não foi muito efetiva em fazer a avaliação e o negócio não foi suficiente, porque, de novo, não está introjetado no DNA. E aí a grande questão é (nem lembro se foi resultado) o tanque de leite deu retorno ou não deu retorno? Ainda, vamos abortar o tanque de leite? Tem que fazer aquilo... Assim, não deu tempo de a gente ver, talvez, mudança de desenho de projeto. E, de novo, eu acho que a avaliação não virou processo. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

Na visão do entrevistado EpR1, o caso do MSF estaria relacionado à uma má utilização da avaliação, devido à falha dos avaliadores na escolha de metodologias e técnicas adequadas, mesmo que não intencional. Esse comentário remete a Alkin e Coyle (1988), que argumentam que a não utilização poderia também ser justificada pela inconsistência da avaliação frente aos seus *stakeholders*. O embate entre os avaliadores e a própria coordenação do processo, somado às mudanças no projeto de avaliação original, levou a ilegitimidade da avaliação frente ao EpR, demandante inicial da proposta.

Na visão do gestor do projeto MSF, ao contrário das percepções dos avaliadores e do entrevistado EpR1, as recomendações da avaliação foram, sim, utilizadas, e o entrevistado ainda cita exemplos de como o estudo retroalimentou o projeto. O entrevistado cita que nem todas as recomendações foram seguidas, e nem sequer ele se lembra de todas. No entanto, três exemplos são reforçados pelo gerente: 1) a implantação de um laudo de viabilidade antes de se implantar um tanque de leite em uma região; 2) a maior vinculação entre a marca do programa e suas ações; 3) o cancelamento da obrigação do associativismo para ingressar no projeto como exemplifica o relato abaixo.

> Um dos apontamentos da avaliação foi a necessidade ou a conveniência de fazer um laudo de viabilidade na parte dos tanques de leite. Hoje, quando a gente vai colocar um tanque de leite eles fazem um laudo de viabilidade se é adequado lá ou não. Isso

> foi implementado durante a pesquisa. Outra sugestão que fizeram, e que nós tentamos corrigir isso, é mandar vincular mais marca do programa. Eles detectaram que muita gente recebia do Minas Sem Fome e não sabia que era o Minas Sem Fome nem que era coisa do governo de Minas, muitos achavam até que era o Fome Zero. Então, esses detalhes de associar a marca do Minas Sem Fome, marca do Governo de Minas ao programa nós trabalhamos mais após a pesquisa. Outra questão que eles colocaram, outro questionamento que eles fizeram é que faz parte do Minas Sem Fome muito a questão do associativismo. E eles detectaram que o associativismo está mais ligado à exigência para o acesso ao programa do que realmente é um beneficio do programa. Então nós começamos a atentar para isso, ou seja, a ideia de associação, o programa tem que ajudar nisso e não para usar isso como um requisito do programa, tem que entrar com uma procuração e não uma exigência legal só. Então tinha que ir mais além do que isso. Eles nos apontaram isso e nós concordamos com eles. Tem mais coisa, mas agora eu não vou lembrar. (GERENTE DO PROJETO MSF)

Observa-se, pelo relato do gestor do projeto, que algumas das recomendações foram seguidas e implementadas no projeto, o que elucida, em um primeiro momento, uma possível utilização do tipo instrumental, caso em que as conclusões das avaliações são utilizadas fundamentalmente para a tomada de decisões ou para a solução de problemas dos programas (LEVITON; HUGHES, 1981). Nota-se, no entanto, que o momento vivido pelo programa na época de conclusão do relatório de avaliação, como relatado pelos avaliadores, não era estável. No período, o projeto apresentava dúvidas quanto ao seu orçamento e implementação.

É possível verificar, assim, que o uso das recomendações foi circunscrito mais aos técnicos dos programas, que ao alto escalão do governo, cujo poder de decisão é limitado. Conclui-se, portanto, pela presença forte do uso conceitual (LEVITON; HUGHES, 1981) na avaliação do MSF. O uso conceitual, de acordo com Weiss (1998), acontece quando, mesmo que os técnicos sejam incapazes de utilizar os resultados das avaliações para a tomada de decisão no momento em que as conclusões das avaliações são apresentadas, as conclusões das avaliações podem modificar o entendimento desses técnicos sobre o que constitui o programa e como ele deveria funcionar. Dessa maneira, o uso conceitual, embora não se tenha uma decisão forte e imediata decorrente de uma avaliação, possibilita a geração de novas ideias e direciona formas diferentes de ação para o programa.

Ademais, na visão dos avaliadores, a avaliação poderia ser mais bem aproveitada e aprofundada em novos estudos, de acordo com os relatos, mas isso não aconteceu. Outro tipo de uso decorrente da avaliação do MSF é o uso acadêmico, para a criação de artigos, desenvolvimento de metodologias e, assim, crescimento dos conhecimentos da categoria de avaliadores. Assim, apesar de não ter sido aproveitada em sua totalidade, os próprios

avaliadores a utilizaram para outros fins como, por exemplo, na confecção de artigos acadêmicos como afirma o avaliador MSF2.

> De percepção, o banco de dados pode ser muito mais explorado, e aí a gente fez uso do nosso lado acadêmico, a gente usou isso pra produção de alguns artigos e rendeu uns três artigos. (AVALIADOR MSF2, grifo da autora da dissertação)

> Inclusive, a gente pensou numa continuidade da avaliação, a gente chegou apensar nessa possibilidade, mas como eu saí do CEPP, acabou que eu acho que isso não foi levado adiante. (AVALIADOR MSF3)

Quando questionado sobre a possibilidade de continuar os estudos avaliatórios do MSF, o gestor do projeto afirmou que seria bem-vindo. Contudo, embora existisse uma vontade por parte do projeto em fazer avaliações periódicas, a justificativa é que os estudos desse tipo são custosos financeiramente e haveria de ter um estímulo por parte da SEPLAG novamente para isso acontecer.

> Qualquer avaliação do programa por fonte externa, ela é muito bem-vinda! O problema é que ela é cara né? Nem sei quanto custa, mas não é fácil fazer. Mas se tivesse condições, e a SEPLAG se interessasse, eu acho que é válido repetir periodicamente. Porque o olhar externo é muito enriquecedor. E um olhar diferente preparada para ação só pode fazer bonito. Só tem medo de avaliação quem não faz em bom tempo o trabalho. Como a gente está querendo fazer bem feito... (GERENTE DO PROJETO MSF, grifo da autora da dissertação)

Nota-se que o gerente do projeto apresenta, assim, uma visão positiva da avaliação, sendo alheio aos problemas de metodologia e falta de comunicação entre EpR e avaliadores. É possível supor também que o uso processual, inerente ao aprendizado dos gestores e técnicos do programa sobre os métodos e habilidade necessárias a uma avaliação (PATTON, 2008), também não foi estimulado, dado a ausência de depoimento dos entrevistados da EMATER sobre esse fato. Uma nova avaliação só aconteceria na instituição novamente, caso a SEPLAG, ou outro órgão da administração do governo, incentivasse como relatado no trecho acima. Contratar uma avaliação por iniciativa própria, como demandantes da avaliação, não é uma hipótese considerada.

Na transição entre os governos, entre 2010 e 2011, e consequente replanejamento da estratégia, mudanças no conceito dos projetos estruturadores ocorreram. Nessa mudança, o projeto MSF deixou de ser estruturador para constituir uma rotina da EMATER. Com essa mudança, o projeto abandona o acompanhamento intensivo realizado pela SEPLAG e também o fluxo orçamentário garantido inerente aos projetos estruturadores, como explicado no

capítulo seis. Na visão do gestor do projeto, essa mudança, de estruturador para associado, não teve nada a ver com a avaliação do projeto. Foi uma mudança conceitual, devido à natureza contínua do projeto, sem um fim preestabelecido.

> E também o Minas Sem Fome, ele é continuo, ele não tem um começo e fim. Por exemplo, um programa não sai daqui e não chega lá, ele é uma ação continuada, então daí a idéia de tirar de estruturador e colocar como associado, porque você não tem uma ideia de projeto. Começou em 2004, então, você não tem uma linha de base. "Ele começou com tantas pessoas, ao final de quatro anos nós vamos estar com tantas pessoas e o resultado vai ser esse". Não tem como. Minas Sem Fome é um programa de fomento e ele visa combater a insegurança alimentar, mas não tem um fim. (GERENTE DO PROJETO MSF)

Como fechamento do capítulo e considerações gerais do processo de avaliação do projeto MSF, destacam-se, assim, características que podem ter influenciado positivamente o uso instrumental da avaliação do MSF como, por exemplo, a boa receptividade da EMATER quanto à avaliação e à cumplicidade entre avaliadores e gestor do projeto. Como aspectos que podem ter influenciado negativamente o uso, enumeram-se: a dificuldade de delimitação do estudo de impacto do projeto que repercutiu em uma estratégia diferenciada da inicialmente negociada com a coordenação do projeto de avaliações no EpR; os problemas de natureza metodológica como, por exemplo, as dificuldades relacionadas à seleção do grupo-controle e grupo tratamento e de resgate dos dados beneficiários para a mensuração do impacto e, por fim, a falha de legitimidade recíproca entre EpR e grupo de avaliadores da FJP, sendo o gestor do projeto alheio à ausência de consenso entre eles.

10 AVALIAÇÃO DO PROJETO ACELERAÇÃO DA APRENDIZAGEM NO NORTE DE MINAS, JEQUITINHONHA, MUCURI E RIO DOCE

A avaliação do projeto estruturador Aceleração da Aprendizagem no Norte de Minas, Jequitinhonha, Mucuri e Rio Doce, também conhecido como Acelerar para Vencer (PAV), constitui uma das avaliações realizadas pelo CEDEPLAR, como explicado no capítulo sexto deste trabalho. Para a pesquisa empírica deste trabalho, contou-se com a participação de cinco entrevistados diretos da avaliação do projeto, o que inclui: gerente do projeto, técnico do EpR que acompanhou o projeto de avaliação, dois componentes da Superintendência de Informações Educacionais, setor responsável pelas avaliações na SEE e representante da coordenação do CEDEPLAR. Também incluem-se percepções de entrevistados da coordenação do EpR.

O capítulo está organizado em quatro seções que remontam o histórico da avaliação do projeto, a partir da percepção dos entrevistados, desde a concepção da avaliação e expectativa inicial, até a sua conclusão, com alusão ao tipo de utilização ocorrido. Na primeira seção apresenta-se o contexto em que a avaliação foi inserida, seguido pela implementação e envolvimento dos atores na execução da avaliação. A terceira e quarta seção tratam, respectivamente, do processamento de informações e da utilização da avaliação.

10.1 Contexto geral da avaliação

O projeto Aceleração da Aprendizagem no Norte de Minas, Jequitinhonha, Mucuri e Rio Doce, também conhecido como Acelerar para Vencer (PAV), iniciou-se em 2007, tendo como proposta combater as taxas de distorção idade-série e promover formas de acelerar os estudos para alunos com atraso escolar, com foco nas regiões do Norte de Minas, Jequitinhonha, Mucuri e Rio Doce. O projeto tem como objetivo "[...] aumentar a proficiência média dos alunos do ensino fundamental e reduzir, progressivamente, a taxa de distorção idade-série das regiões Norte de Minas, Jequitinhonha/Mucuri e Rio Doce" (MINAS GERAIS, 2009). Trata-se, assim, de um projeto que visa desenvolver uma metodologia de ensino específica para a aceleração da aprendizagem de alunos cuja distorção idade série seja superior a dois anos letivos; elaborar recursos didáticos adequados a esse público e capacitar os professores para adoção dessa metodologia.

O PAV visa, assim, a correção da distorção idade-série entre alunos com dois ou mais anos de defasagem, matriculados no ensino fundamental (do terceiro ao oitavo ano), sem limite de idade, da rede municipal e estadual de ensino, de municípios do Norte de Minas, Jequitinhonha, Rio Doce e Mucuri, em Minas Gerais. As escolas poderiam se inscrever no PAV, desde que tivessem pelo menos 20 alunos com distorção idade-série igual ou superior a dois anos, além de espaço físico para a formação de turmas especiais. Serão formadas turmas especiais para os alunos participantes do projeto, que recebem recursos didáticos específicos e terão atendimento de professores treinados e capacitados pela Secretaria do Estado de Educação. Espera-se que, ao final de um ano calendário, os alunos consigam aumentar sua escolaridade, em média, o equivalente a dois anos de estudo, além de melhorar sua proficiência nos estudos e se alfabetizarem, caso não consigam ler e escrever. De acordo com o gerente do projeto PAV,

> o PAV é um projeto que faz parte do esforço do Governo para elevar os índices socioeconômicos educacionais do grande Norte de Minas. Então, o PAV é um projeto que foi implantado e é um projeto estruturador para reduzir as taxas de distorção e de antiescolaridade nessa região que eram as mais altas do Estado. O PAV é um projeto que precisa muito ter professor com o perfil adequado, porque são os meninos que perderam o bonde da escola, eles perderam ou porque foram embora, saíram da escola, ou porque não conseguiram aprender porque nós não conseguimos ensinar. Então, são meninos que a gente considera sofridos, são meninos que, por eles, estavam indo embora, porque com muita insistência da gente né. Então, uma das grandes linhas do PAV é levantar a autoestima, resgatar a autoestima dos meninos, porque para as escolas, muitas delas, eles são aqueles que não querem nada, e não é bem verdade. Quando muito o projeto começa num grupo pequeno, de um ano, e depois vai para todas as outras. O PAV foi assim. O PAV começou só no grande norte, porque ele era estruturador no grande norte, acabou que nós levamos ele para o Estado inteiro. Por quê? Porque não era só no norte que tinha altas taxas de distorção, as outras regiões do estado também tinham. Então em 2008 nós ficamos só no norte, mas 09, 10, 11 e 12 no Estado inteiro. (GERENTE DO PROJETO PAV)

Para os integrantes da Superintendência de Informações Educacionais, quem recebeu e coordenou a avaliação dentro do SEE, a pesquisa foi mais uma demanda do EpR que da própria educação. Isso devido ao interesse e conhecimento das pessoas que já trabalhavam no EpR à época que, de acordo o entrevistado SEE2, já detinham a *expertise* necessária ao desenvolvimento de avaliações de políticas públicas, preparação essa que não estaria presente em todas as secretarias e pessoas que trabalhavam no governo. A carteira de projetos que seria avaliada também chegou pronta, com a decisão de avaliar, além do PAV, os projetos estruturadores Escola em Tempo Integral (PROETI) e o Programa de Ensino Médio Profissionalizante, mais conhecido pela sigla PEP. O quarto projeto estruturador avaliado, o

projeto Poupança Jovem, apesar de ser um projeto que trabalha com os alunos das escolas, estava inserido na carteira de projetos da Secretaria de Estado de Desenvolvimento Social (SEDESE). Os entrevistados concordam que a demanda de avaliação não partiu da SEE.

> A pesquisa, na verdade, veio do Estado para Resultados. Na verdade, o Estado para Resultados contratou o CEDEPLAR. Então, eu não sei se a demanda realmente foi da Educação de falar "Olha, vamos fazer uma avaliação de impacto para ver se essas políticas estão dando certo" ou se realmente foi uma demanda. Eu acredito que foi uma demanda do Estado para Resultados até por causa do *staff* que tinha lá, pessoal com conhecimento na área, com certa preocupação. Já chegou: "Olha, nós vamos fazer uma avaliação de impacto em três projetos da secretaria e, mais um que não é a secretaria, o Poupança Jovem". (EQUIPE SEE2, grifo da autora da dissertação)

> Na realidade, apesar de eu achar que é superimportante uma avaliação de impacto, não saiu da secretaria. Saiu uma política que estava tendo no Estado, no Escritório e na Fundação João Pinheiro, que queria avaliar alguns projetos, as áreas mais importantes que seria, no caso, a educação. Então, não foi uma demanda da secretaria. (EQUIPE SEE1, grifo da autora da dissertação)

O contrato para a realização das avaliações dos projetos educacionais, no entanto, não partiu do mesmo contrato com o Banco Mundial que contava com a execução da FJP. As quatro avaliações, como afirmado anteriormente no capítulo sexto, foram executadas pelo CEDEPLAR, por meio do contrato celebrado entre SEPLAG e FUNDEP – contrato nº631/08. Constituiu objeto do contrato a prestação de serviços técnicos de consultoria envolvendo a avaliação de impacto de projetos estruturadores inseridos na área da educação e a avaliação de indicadores finalísticos, com treinamento na utilização de bases de dados específicas. O entrevistado EpR3 explica que o contrato das avaliações era com contrato *guarda-chuva*, sendo aplicados valores específicos para a avaliação e para a prestação de consultoria sobre indicadores educacionais, sendo os dois detalhados por termos de referência posteriores.

Uma especificidade do contrato firmado entre o governo e o CEDEPLAR trata da existência de uma cláusula de sigilo acerca dos conhecimentos e informações técnicas advindas das avaliações dos projetos educacionais. Nesse sentido, na tentativa de manter o caráter confidencial da pesquisa avaliatória do PAV, os resultados da sua avaliação não serão aqui apresentados, como ocorreu nos demais projetos avaliados, como o Minas Sem Fome, o Viva Vida e o Gestão Integrada. Objetiva-se, portanto, manter a confiabilidade do relatório do PAV *a priori* não será citado nenhum trecho do relatório final de avaliação, salvo os comentários advindos das entrevistas dos atores envolvidos que, de alguma forma, comentam alguns dos resultados que lhes vêm à memória.

Quanto à cláusula de confiabilidade das avaliações educacionais, cabem algumas considerações sobre a publicidade das avaliações de políticas públicas. Faria (2005) argumenta que sistemas institucionalizados de avaliação de programas e políticas públicas constituem elementos profícuos na criação e no aperfeiçoamento de mecanismos de prestação de contas do desempenho das burocracias e representantes políticos. As avaliações compõem, assim, peças válidas do mecanismo de prestação de contas de um governo, na tentativa de promover níveis crescentes de responsabilização, verificação e transparência, ou seja, como elemento central de *accoutability*, dependendo do grau de difusão da informação gerado pelas avaliações. A criação de cláusulas de confiabilidade para a apresentação de resultados de qualquer avaliação contra esse princípio de transparência das informações geradas pelas avaliações e da posição dos estudos avaliatórios como instrumentos de prestação de contas. Ainda, quanto à interposição de normas de sigilo para as avaliações, parece por demais óbvio excluir a presença do caráter eminentemente político da avaliação. A obrigação da confiabilidade dos dados das avaliações educacionais suscita o receio, por parte de seus dirigentes, de possíveis resultados negativos que poderiam interromper as políticas ou mesmo promover propagandas contraproducentes contra o governo que criara a política. Propaganda esta que poderia ser utilizada, até mesmo, pela oposição. Como exemplo da falta de transparência dos resultados, o entrevistado do EpR comenta como os resultados das avaliações educacionais, que poderiam ser divulgados e utilizados em outras searas, como a área acadêmica, não foram disseminados. O motivo da ausência de transparência dos relatórios, no entanto, é desconhecido pelo entrevistado, que aposta no palpite de que os resultados não viraram notícia, justamente por serem inconclusivos.

> A disseminação acabava por aí, nenhuma das avaliações foi colocada como disponível na internet, os bancos de dados do Governo nunca foram disponibilizados para pesquisadores ou para estudantes, os resultados da avaliação também não foram, ainda que inconclusivos, ou talvez justamente por isso, não foram propagandeados, não saiu na comunicação do Governo nem na comunicação externa também. Ficou na gaveta. (TÉCNICO EPR3)

De acordo com a coordenação do CEDEPLAR, as avaliações dos projetos educacionais não foi o primeiro trabalho desenvolvido para o governo. A primeira iniciativa de contratação do centro de pesquisas foi para a realização de um diagnóstico de possíveis desenhos de políticas públicas, com marco teórico definido, ciclo de vida e proposta posterior de avaliações para aquelas propostas. No entanto, para o entrevistado, a experiência anterior foi um pouco

frustrante por não ter sido aplicado no desenho de políticas de fato, como se relata neste primeiro trecho da entrevista.

> Porque, na verdade, seguinte, o primeiro relatório que a gente fez para o Estado para Resultados foi em 2006, 2007. O primeiro relatório era para desenhar um conjunto de políticas públicas. "Desenhar", então não tinha avaliação, tinha até uma sugestão de avaliação, de como avaliá-los, mas o ponto central era o desenho de uma série de políticas. Esse relatório, que foi de 2006, tinha dezenas de políticas ali desenvolvidas no contexto do marco teórico, do ciclo de vida etc, etc, isso nunca foi usado. Zero. Foram as primeiras experiências. Foi um pouco frustrante nesse sentido. Não houve nem discussão. (COORDENAÇÃO CEDEPLAR)

E de fato, não somente os estudos avaliatórios correm o risco de não serem utilizados para a melhoria dos projetos avaliados, mas também qualquer outro tipo de evidências, como o exemplo do estudo de possíveis projetos para a implantação, que se refere a um tipo de avaliação *ex-ante* de políticas. Nas palavras do entrevistado, ele não consegue visualizar alguma política do governo que se assemelhe a alguma política sugerida no diagnóstico. Entretanto, a iniciativa, apesar de não utilizada, ensejou a realização da avaliação do projeto Poupança Jovem e posteriormente, as avaliações dos demais projetos da SEE, entre eles o PAV. De acordo com o entrevistado, da mesma forma que a carteira de projetos a ser avaliada já chegou definida na SEE, já chegou também determinada no CEDEPLAR, com a deliberação dos projetos que seriam avaliados.

> Logo, veio a demanda para a gente avaliar o Poupança Jovem, aí teve toda a avaliação do Poupança Jovem. No meio, o Secretário do Estado para Resultados foi encomendando novas avaliações de outros programas da educação. Não sei exatamente como é que ele escolheu o que ia ser avaliado, tá? Ele falou assim: "tem esse projeto aqui, tem esse projeto aqui, vocês querem avaliar?" Então, tinha o PEP, um que chamava PAV (Avançar para Vencer) e o Escola Tempo Integral. Então, nós avaliamos esses quatro. E aí a experiência interessante. (COORDENAÇÃO CEDEPLAR)

As expectativas iniciais do centro de pesquisas quanto às avaliações dos projetos educacionais eram elevadas, com a esperança de se criar uma base de dados permanente para avaliações de políticas públicas independentemente dos projetos avaliados. De acordo com o entrevistado, a ideia seria estabelecer uma parceria entre governo e universidade para a criação de um sistema maior de avaliações a exemplo de outros países no mundo. Tais expectativas iriam além do uso das avaliações, mas igualmente do estabelecimento de organizações voltadas especificamente para a realização de avaliações constantes e periódicas das ações do governo, remetendo ao tipo de uso *enlightenment* (WEISS, 1998) no caso em que as avaliações são utilizadas para influenciar outras instituições ou eventos que vão além do escopo de

determinado programa avaliado, ultrapassando seus limites originais e pautando, não somente os coordenadores de programas avaliados, mas também uma determinada comunidade acadêmica, como o CEDEPLAR, por exemplo, e coalizões políticas interessadas.

> Porque tinha uma ideia. No início, uma coisa que culminou a partir daí, uma ideia que surgiu, mas que, infelizmente, também não foi para frente, era a de, exatamente, montar dentro do Estado ou aqui na universidade junto com o Estado um centro de avaliação que a alicerce começaria com a construção de uma grande base de dados. Criar uma base permanente, independente dos projetos, uma base de dados que a gente acompanhasse os indivíduos. O programa entra, mas a gente sabe como é que o individuo está indo na escola. Então, a gente teria uma base permanente de avaliação. Isso é uma coisa que tem na Inglaterra, que tem em alguns lugares do mundo. (COORDENAÇÃO CEDEPLAR)

Na perspectiva do entrevistado do EpR, as expectativas eram *muito altas*. De acordo com a sua percepção, a esperança seria de que os relatórios de avaliação gerassem informações e recomendações e estas, por sua vez, guiariam os próximos passos do governo. Tal percepção remete ao uso instrumental, dimensão de uso que diz respeito aos casos nos quais as conclusões das avaliações são utilizadas fundamentalmente para a tomada de decisões ou para a solução de problemas dos programas (LEVITON; HUGHES, 1981).

> Minhas expectativas eram muito altas, eu tinha expectativas muito altas de que daria bons resultados e que daria informações e recomendações de quais seriam os próximos passos. (TÉCNICO EPR3, grifo da autora da dissertação)

De acordo com o gerente do projeto, outros tipos de avaliações do projeto já apontavam para um resultado positivo dos impactos do PAV, portanto, a expectativa com a avaliação seria mostrar efeitos de redução das taxas de distorção idade-série nos alunos que participavam do projeto. O entrevistado chega a falar que a equipe do projeto tinha *certeza* que os resultados seriam positivos e, depois, corrige o relato dizendo que era uma *expectativa* da equipe. Além disso, na fala, a percepção é de que a avaliação se esgota em si mesma, não havendo indicações sobre a importância das conclusões da avaliação para melhoria e aperfeiçoamento do projeto.

> Pelas análises que a gente vinha fazendo não pela avaliação de impacto, mas avaliação externa da secretaria que os alunos do PAV também fazem pelo PROEB, que era o 5° e 9° ano, pelos resultados da avaliação externa que a gente estava acompanhando, pelos resultados dos censos escolares, que a gente estava vendo que estávamos reduzindo as taxas de distorção, a gente tinha certeza, pelo menos a nossa expectativa é de que essa avaliação de impacto iria mostrar que os resultados seriam positivos. (GERENTE DO PROJETO PAV, grifo da autora da dissertação)

Outro ponto que chama a atenção a partir dese relato do gerente do projeto é a existência de informações concorrentes no Sistema Mineiro de Avaliação da Educação Pública (SIMAVE) que, de alguma forma, poderiam competir com os resultados da avaliação de impacto. Implantado em 2003, o SIMAVE é responsável pelo desenvolvimento das avaliações do sistema educacional com foco no ensino do sistema escolar e, não no aluno. Nesse sentido, o Projeto propõe atividades relativas à realização de duas avaliações fundamentais: o Programa de Avaliação da Rede Pública de Educação Básica (PROEB) e o Programa de Avaliação da Alfabetização (PROALFA). Em linhas gerais, o PROEB, realizado desde 2000, tem por objetivo avaliar as escolas da rede pública, no que concerne às habilidades e competências desenvolvidas em língua portuguesa e matemática dos alunos que se encontram no 5° ano e 9° ano do ensino fundamental e 3° ano do ensino médio. De forma semelhante, o PROALFA, cuja primeira avaliação ocorreu em 2005, verifica os níveis de alfabetização alcançados pelos alunos do 3° ano e 4° ano do ensino fundamental da rede pública. Na visão de Cousins e Leithwood (1986), quando confrontados com um problema, os tomadores de decisões confiam em várias fontes de informações diferenciadas, além dos resultados das avaliações e, por isso, consideraram a variável informação competitiva, que constataram fontes alternativas de informação competindo com os resultados da avaliação para tomada de decisões, uma questão importante para a sua utilização. No caso da SEE, quando o próprio gerente do projeto e outras pessoas envolvidas confiam mais em outras fontes de informação que na própria avaliação, como o PROEB, esta se mostra concorrente poderosa com os resultados das avaliações. Os relatos abaixo comprovam a existência das avaliações que poderiam cumprir com o resultado das avaliações de impacto e, até, a preferência, pelo gerente do projeto, por aquele tipo de avaliação.

> A secretaria é o SIMAVE - o Sistema Mineiro de Avaliação da Educação, que tem o PROALFA, que avalia a alfabetização, que tem o PROEB, que avalia o 5° ano, o 9° ano e o 3° ano do ensino médio. E tem o IDEB que, de certo, não é uma avaliação, mas é o índice de desenvolvimento que leva em conta não só as taxas de distorção de repetência, mas leva em conta também o desempenho dos meninos na Prova Brasil. Então a gente acha que essas duas avaliações nos dá também medida de que se nossas ações implementadas nas escolas estão indo bem. Por exemplo, os anos iniciais, os anos finais. Então o ensino fundamental está bem comparativamente com o Brasil, embora para nós nunca está bom porque nós queremos melhorar cada vez mais. Agora, uma coisa que eu queria dizer que, embora a gente considere a avaliação de impacto muito importante, mas eu acho mais difícil de ser feita na educação. Uma coisa que eu acho que avalia muito bem também nosso trabalho é a avaliação externa que nós fazemos. Porque, no fim, o que que a gente quer com qualquer projeto, com qualquer ação, com qualquer programa? É que os meninos tenham um bom resultado de aprendizagem, e a avaliação externa eu acho que nos dá essa resposta. (GERENTE DO PROJETO PAV, grifo da autora da dissertação)

E, de fato, o entendimento inicial dos gerentes de projetos quanto à relevância das avaliações de impacto contratadas não foi algo generalizado a princípio de acordo com as entrevistas. Para o entrevistado SEE1, as avaliações foram recebidas como uma atividade secundária na SEE, *não seria uma coisa tão importante*. Já o entrevistado do CEDEPLAR explica que, na verdade, os gestores dos projetos avaliados ficaram *temerosos* com a avaliação, na medida em também veem o seu trabalho avaliado. Em outras palavras, a qualidade dos serviços públicos é um desafio também de seus gestores, que detêm a capacidade de, se bem administrados, originar produtos aptos a satisfazer as necessidades dos beneficiários, como afirma Pisco (2006). Assim, esses mesmos gestores, que se dedicam com empenho para que as ações do projeto saiam como planejado, sentem-se fracassados se seu projeto não alcança impacto quando avaliados, como se o trabalho dos próprios gestores estivesse passando pela avaliação. Contudo, como explica o entrevistado no segundo trecho abaixo, ainda que, a administração do projeto seja perfeita, os gestores teriam que entender que se o projeto não tiver os incentivos e o desenho correto, nunca alcançará impacto, independentemente do tipo de gerenciamento.

> Foram três gerentes, porque eram três projetos, alguns ficaram um pouco receosos. No final foi um pouco... acho que eles enxergaram a importância da avaliação, mas não foi a princípio não. No primeiro momento eles tiveram uma, a palavra não é "resistência", mas acharam que não seria uma coisa tão importante. (EQUIPE SEE1)

> Agora a gente está conversando com o gestor lá, a primeira coisa dos sujeitos é que eles ficam muito temerosos. Fica. Porque o cara está gerindo uma administração do programa, e a gente está avaliando é o impacto social do programa, aí se a gente não acha impacto "Ah, mas eu estou gerindo ele tão bem". Verdade, mas aí que é o desenho do programa, porque o programa não tem os incentivos corretos. Então, assim, essa é uma dificuldade, então, chega uma hora que enche a paciência do gestor. A gente chega até no gestor maior lá, o cara fala "Ah...". Isso demora, não é uma coisa que você tem imediatamente, a não ser que a gente tivesse essa ideia de uma base de dados permanente e tal. (COORDENAÇÃO CEDEPLAR, grifo da autora da dissertação)

E como os gerentes dos projetos avaliados não *abraçaram* a causa das avaliações de impacto educacionais, a interface entre avaliadores e SEE, e EpR e SEE aconteceu com a Superintendência de Informações Educacionais, uma unidade voltada para o acompanhamento de projetos e efetivação do SIMAVE. Para os integrantes da superintendência (entrevistados equipe SEE1 e equipe SEE2), os estudos avaliatórios eram uma iniciativa do EpR e, portanto, coordenados pelo EpR. A função da superintendência seria fazer a intermediação entre EpR, os gerentes de projetos e o CEDEPLAR; fornecer dados e auxiliar no desenho com os dados.

> O que chegou para a gente: "Olha, vai ter três avaliações de impacto e eu quero que vocês façam essa intermediação", mas sempre quem coordenou foi o Estado para Resultados, a gente nunca coordenou essa avaliação. Até porque não saiu nada da Secretaria de Educação para pagar essa pesquisa, foi recurso do Estado para Resultados. Aí chegou lá e a gente foi tocando. (EQUIPE SEE2, grifo da autora da dissertação)

> Na verdade, a avaliação de impacto do PAV, juntamente com o Tempo Integral e o PEP, foi uma iniciativa não da secretaria, mas na época era do Estado para Resultados. Como lá na Superintendência de Informações Educacionais você tinha todas as bases de dados tanto do censo com as avaliações e também uma diretoria ligada que era de acompanhamento de projetos, a superintendência foi o elo entre os escritórios e os gerentes, a secretaria, e o CEDEPLAR, que foi quem fez essas três avaliações. Então o papel foi, principalmente, em questão de fornecer os dados, todo o projeto, a parte metodológica que o pessoal do CEDEPLAR juntamente com o pessoal do escritório sempre mandava para a gente para a gente dar uma olhada. Mais ou menos o papel foi esse, principalmente fornecer os dados e auxiliar no desenho com os dados. (EQUIPE SEE1)

Entretanto, na visão do técnico do EpR, mesmo com o acompanhamento da Superintendência de Informações Educacionais, as avaliações eram vistas como uma iniciativa do EpR. Para Bamberger, Rugh e Mabry (2011), embora o desenho de uma avaliação devesse equilibrar diferentes interesses dos *stakehoders* envolvidos, alguns atores detêm maior influência e estão em uma posição mais forte para definir critérios e o próprio foco da avaliação. Os atores alocados no segundo plano, por estarem mais distantes das decisões sobre a condução da avaliação, são mais difíceis de serem incorporados ao processo, como exemplifica o entrevistado.

> Acho que a Superintendência de Informações Educacionais acompanhava, mas também era de uma forma muito solta, e a diretoria de acompanhamento de projetos também. Acho que havia uma sensação dentro da Secretaria de Educação, na época eu não tinha muito acesso ao fluxo lá dentro não, a minha impressão é de que dentro da Secretaria de Educação as avaliações de políticas de educação eram vistas como iniciativas e ações do Estado para Resultados e não do Governo, do Estado, muito menos da Secretaria de Educação. (TÉCNICO EPR3, grifo da autora da dissertação)

Dessa forma, o contexto inicial de introduções das avaliações de impacto educacionais foi marcado pela falta de comprometimento e envolvimento dos próprios gerentes de projetos da SEE quanto à importância das avaliações. De acordo com Cousins e Leithwood (1986), um maior envolvimento, tanto dos decisores, quanto dos próprios gestores, contribui para altos níveis de utilização, bem como alcance de uma melhor aprendizagem sobre o projeto e aumento de fatores tais como comunicação, processamento de informações e credibilidade. Para os autores, atitudes dos gestores em relação à avaliação têm uma relação inequívoca com

a utilização positiva (COUSINS; LEITHWOOD, 1986). A seção seguinte trata do planejamento e execução da implementação da avaliação de impacto do PAV.

10.2 Implementação do processo de avaliação e envolvimento dos atores

O processo de planejamento da avaliação de impacto do PAV, assim como a sua execução, contou com pouca ou nenhuma interação com a equipe gestora do projeto. A participação do gerente do projeto, de acordo com trechos da entrevista, principiou já na fase final da avaliação, na etapa de conclusão do relatório de avaliação, com a análise do relatório e processo de revisão do mesmo. Observa-se, assim, que o contato do gerente do projeto com a avaliação foi feito em dois momentos: no primeiro, quando foi informado da realização da avaliação, já definida pelo EpR; e, no segundo momento, somente quando a avaliação foi concluída, quando os resultados já estavam processados e devidamente organizados no relatório final. O processo de planejamento não contou com qualquer participação do gerente, como relata no trecho abaixo.

> Tivemos umas duas ou três reuniões, mas não para planejar a avaliação de impacto, mais para recebermos o documento, avaliarmos e analisarmos o documento, apresentarmos nossas questões que a gente não tivesse entendendo. Para planejar a avaliação não, foi feita por eles. (GERENTE DO PROJETO PAV)

> A comunicação era bem rara, pode se até dizer uma das falhas da avaliação. (TÉCNICO EPR3)

Os técnicos da Superintendência de Informações Educacionais também participaram pouco do processo de planejamento da avaliação do PAV, embora tenham participado mais que o gerente do projeto. De acordo com o entrevistado da equipe SEE2, a intermediação era sempre realizada pelo EpR em reuniões com os avaliadores do CEDEPLAR, de acordo com um cronograma previamente estipulado no contrato entre as instituições. Não havia encontros não programados, nem um acompanhamento constante. Para o técnico do EpR3, o EpR trabalhava como um tradutor das demandas da avaliadores para a SEE e prioridades da SEE para os avaliadores, com certa autonomia. Já, na percepção do entrevistado equipe SEE2, o contato com a secretaria só era realizado quando os avaliadores necessitavam de dados e informações educacionais específicas, sem nem explicar a necessidade daqueles dados e o motivo pelo qual seriam utilizados; a função da SEE seria, assim, mais procedimental, de atender as demandas sem muito questionamento. Desse modo, enquanto na percepção dos

técnicos do EpR, a instituição funcionava como um tradutor, na opinião dos técnicos da educação envolvidos na avaliação, o EpR ocupava o papel de demandante.

> Os encontros do Estado para Resultados com o CEDEPLAR era mediante um cronograma inicial já previsto, um plano de trabalho previsto no contrato, e os termos de referência já estabeleciam prazos também, meses, eu acho, para o cumprimento. Então bem na época de estabelecer o termo de referência havia o contato e na época de receber o produto havia o segundo contato, não havia um acompanhamento constante e nem... de tempos em tempos, planejadamente. Não tinham encontros intermediários programados não. O Estado para Resultados trabalhava como um intermediário, traduzindo o que era demanda para as políticas de educação para os avaliadores realizarem, e o que eram as respostas da instituição avaliadora para a Secretaria de Educação, mas com muita autonomia. (TÉCNICO EPR3)

> Era sempre intermediado pelo Estado para Resultados. Eles faziam reuniões, vinha o pessoal do CEDEPLAR e a gente ia lá. Geralmente o gerente de projeto nem ia, ia mais a gente. Geralmente era para demandar alguma coisa: "Olha, a gente está precisando disso, base de dados...". Não dava muita informação não. (EQUIPE SEE2)

Como a colaboração entre avaliadores e equipes dos projetos avaliados foi restrita, o aprendizado e a absorção de conhecimento por parte da SEE, quanto ao tema e às práticas de avaliação, foram limitados. De acordo com relatos da equipe da Superintendência de Informações Educacionais, a participação da SEE era de apoio ao CEDEPLAR como um simples fornecedor de dados. As reuniões que contaram com a participação da SEE eram rápidas e muito técnicas, sem estímulos para que os avaliados se sentissem incluídos no processo. De acordo com o relato abaixo, a percepção é de que não havia incentivos para facilitar o entendimento das abordagens avaliatórias pela equipe dos projetos e, assim, da maneira que a avaliação foi conduzida, não houve qualquer assimilação de conhecimento por parte da secretaria de educação, como atenta o entrevistado.

> Na verdade, a participação nossa era só assim, auxiliar a equipe do CEDEPLAR. Assim, até uma crítica que eu faço: primeiro, não teve nenhuma absorção de conhecimento por parte da secretaria em relação de aprender, pelo menos... O contato nosso era sempre assim: "Ah, passa base de dados para a gente... No projeto Tempo Integral está precisando de cadastro, vocês viabilizam isso para a gente?". E a gente fazia. E as reuniões que tinham eram justamente assim, tinham algumas reuniões geralmente com o Estado para Resultados, a gente participava, mas era uma reunião muito técnica e discutia muito rápido também quase todos os projetos da carteira que eles estavam avaliando. (EQUIPE SEE2)

Dentre os motivos que dificultavam a inserção dos gestores da Secretaria de Educação no processo de avaliação dos projetos educacionais, duas questões distintas, comentadas pelo entrevistado do CEDEPLAR, podem ter influenciado negativamente as avaliações: primeiro,

a postura dos funcionários do EpR frente à Secretaria e ao tema e, segundo, as diferenças entre a dinâmica de trabalho do governo e da universidade. Para o entrevistado, a postura formal dos representantes do EpR, que sugeriam mais uma estilo de consultor e, não, de parceiro, ajudava a acirrar o clima de desconfiança entre os gestores da SEE, o que impedia a aceitação da avaliação. Nesse sentido, a avaliação poderia ser atendida com um caráter fiscalizatório, como afirma Patton (2008), e ser confundida com uma auditoria. O comentário do entrevistado exemplifica a situação de que o representante do EpR parecia um *enviado do governador*.

> Um pouco essa coisa da desconfiança dos gestores; inicialmente eles ficam desconfiados e isso impede um pouco a aceitação da avaliação. Até brincava com o cara lá do EpR, ele chegava na Secretaria de Educação, lá na Gameleira, chegava de terno e gravata e eu: "Bicho, você está parecendo um consultor da Price". O cara chegava e os gestores já ficavam com medo, como se ele fosse um enviado do Governador. (COORDENAÇÃO CEDEPLAR)

Quanto à dinâmica de funcionamento entre universidade *versus* governo, é possível verificar diferenças entre os valores da academia e do setor público que poderiam influenciar a avaliação de impacto, fato confirmado por Leviton e Hughes (1981). Para os autores, a diferença de estilos entre os administradores e os pesquisadores também pode afetar a utilização dado que existe pouco contato entre os administradores e a comunidade acadêmica (LEVITON; HUGHES, 1981). No comentário do entrevistado do CEDEPLAR, as avaliações contratadas pelo governo eram vistas como apenas mais uma das pesquisas dentre as variadas atividades presentes na rotina dos pesquisadores, que também se dedicavam às aulas, às publicações e às orientações, como sugerido na fala abaixo. Na visão do entrevistado, o conhecimento de como executar as avaliações deveria ser transferido ao governo.

> Agora, é muito complicado pelo seguinte, trabalhar com universidade também não é simples. Porque seguinte: essa relação é sempre muito difícil. Eu explico. O Estado contrata a universidade, digamos. Muito bem. Vai ter o EpR de um lado e vai ter do outro lado um bando de professor e aluno. Essa é uma relação difícil porque, por exemplo, a gente não tem tempo para ser consultor de luxo e nem estava proposto no negócio, então era bem difícil, porque a gente dá aula, porque a gente é dedicação exclusiva, sou professor, sou orientador, tem que publicar, mais que suficiente para ter meu dia totalmente preenchido. O que o Estado podia ter feito era ir identificando essa coisa da avaliação em outras áreas para que a gente pudesse ir transferindo inteligência. (COORDENAÇÃO CEDEPLAR)

A falta de conhecimento dos gerentes dos projetos educacionais avaliados sobre o tema das avaliações de impacto, devido a sua especificidade e complexidade metodológica, foi comentada por diferentes atores, nas entrevistas. O foco na causalidade, marca das avaliações

de impacto, exige a utilização de determinadas metodologias (KHANDKER; KOOLWAL; SAMAD, 2010). Para estimar o efeito causal de um programa sobre seu público-alvo, o método escolhido deve simular o que aconteceria com os beneficiários caso o programa não tivesse ocorrido, com o objetivo de descobrir quais as mudanças nos beneficiários são diretamente atribuíveis ao programa em questão. Assim, de forma simplificada, para auferir o impacto de uma intervenção, utilizam-se diferentes métodos capazes de comparar os efeitos esperados do programa em um grupo de tratamento (grupos de indivíduos que participaram do programa) e outro grupo de controle (grupo de indivíduos com características semelhantes ao grupo tratamento, mas que não participaram do programa) em relação aos resultados/mudanças ao longo do tempo observados. No entanto, de acordo com os entrevistados, essa dinâmica das avaliações de impacto não era algo trivial para um grupo de pedagogos que gerenciavam os projetos avaliados na SEE, principalmente no que tange `a diferenciação entre os grupos-controle e tratamento. Os entrevistados da SEE justificam, assim, a não participação quase que natural dos gerentes da Secretaria pela ausência de conhecimento sobre as especificidades de uma avaliação de impacto. Não havia como demonstrar interesse e participar das avaliações se não havia qualquer vínculo entre avaliadores e avaliados e se não existiam incentivos para incluir os gerentes no processo de avaliação, como relatam os entrevistados abaixo.

> Quando fez o desenho, acaba que na avaliação de impacto há uma parte metodológica muito forte né. E os gerentes não têm muito entendimento em relação a isso. Então, assim, eu não sei se eles acompanharam, conseguiram entender a parte metodológica, principalmente na definição do grupo controle de tratamento, porque lá na Secretaria de Educação eles são pedagogos. Área de estatística não é o forte lá, e não teria que ser, porque eles são pedagogos. Então, assim, teve um pouco de falta de entendimento de como foi definido toda a técnica mesmo, a parte metodológica. (EQUIPE SEE1)

> Na verdade, seguinte, os gerentes dos projetos participaram pouco da avaliação, mas não tiveram tanto interesse eu acho até pela dificuldade do tema. Porque, quando eles iam à reunião, geralmente era discutido metodologia, e é uma coisa complicada principalmente para pessoal da área pedagógica. Acabou que eles não participaram e quem tomou a frente disso mesmo foi nosso setor lá [Superintendência de Informações Educacionais]. (EQUIPE SEE2, grifo da autora da dissertação)

O próprio gerente do projeto confessa que só foi conhecer a teoria das avaliações de impacto após a conclusão da avaliação do PAV. Como pode ser verificado no trecho abaixo, o gerente relata que, na época em que foi realizada a avaliação, um estudo avaliatório de mensuração de impacto era como as demais avaliações. Somente após a participação em um curso desvinculado do movimento de avaliações coordenado pelo EpR que foi conhecer as

exigências e meticulosidade das avaliações de impacto, mas, nesse momento, a avaliação do PAV já havia sido concluída.

> Na verdade, a gente não tinha... Pelo menos, eu, pessoalmente, eu não tinha muita experiência nessa questão de avaliação de impacto. Eu fiz um curso no Rio de Janeiro que até a Secretaria de Educação que nos mandou. Todos nós, gerentes, fizemos o curso de avaliação de impacto na secretaria. A época de 2007 e 2008, 2008 ou 2009 – não me lembro muito bem, que foi feita a avaliação de impacto, eu não tinha muita experiência nessa área. Depois que a gente aprendeu. Fez esse curso lá que foram 40 horas, a semana toda, então, a gente viu a importância dessa avaliação e como ela é minuciosa e como ela é exigente, é muito exigente na sua metodologia. Mas na época eu não tinha muito conhecimento, para mim era uma avaliação como as demais. (GERENTE DO PROJETO PAV)

O papel do Banco Mundial, como treinamento dos envolvidos nas avaliações da FJP, foi um pouco diferenciado nas avaliações educacionais. Os avaliadores dos projetos da educação, até pela experiência que o próprio CEDEPLAR já detinha, não participaram dos treinamentos e discussão da fase inicial de avaliações com o Banco Mundial. Na percepção da coordenação do CEDEPLAR, de acordo com os relatos, o papel da instituição internacional foi importante como propulsor do modelo de avaliações, como incentivador de um *clima de avaliação*, clima esse que desapareceu no governo posteriormente.

> Tinha o Banco Mundial, só que a remuneração da universidade aqui para a contratação não era feita pelo Banco mundial não. O Banco Mundial, o que ele fez foi incentivar o Estado a fazer. Ele veio e treinou o pessoal da Fundação. Mas o que o Banco Mundial fazia era incentivar o Estado a avaliar, aí tinha reunião e então isso tudo gerava um clima de avaliação e que não tem mais não. (COORDENAÇÃO CEDEPLAR)

Além da ausência de incentivos para o engajamento dos profissionais da Secretaria de Educação, outro ponto ressaltado nas entrevistas foi a falta de tempo dos gestores dos projetos. Assim como na Secretaria de Saúde, os gerentes e equipes do projeto se mostraram presos `a rotina confusa dos projetos e sem tempo para se dedicar à avaliação e monitoramento dos projetos. O gerente do PAV, por exemplo, além de se dedicar às atividades do projeto, também exercia o cargo de gerência de outro projeto da Secretaria: o Programa de Intervenção Pedagógica (PIP). Dado o acúmulo de funções, somado à falta de conhecimentos sobre avaliações e ausência de incentivos para participar, as questões mais técnicas da avaliação não eram sua prioridade, como relatado nesta entrevista.

> É porque eu era gerente de dois projetos enormes, e o PIP me absorvia nessa época mais ainda. O PAV começou em 2008, o PIP 2007 e 2008, então, essas questões, vamos dizer assim, mais técnicas, e não tanto pedagógicas, ficavam mais com o

> secretário adjunto, com a equipe de avaliação e menos com a gente. A gente recebia mais os resultados para análise. (GERENTE DO PROJETO PAV)

Outro ponto que merece destaque são as características organizacionais e políticas da instituição que, de alguma forma, dificultam a implantação das avaliações de impacto e, assim, a utilização das mesmas como correção de rota. Para o gerente do projeto, um dos motivos que atrapalham a realização de avaliações de impacto na Secretaria refere-se ao *trade off* entre extensão para tempo de execução da pesquisa e a necessidade urgente de expansão de projetos educacionais. Em outras palavras, como o desenvolvimento de uma avaliação de impacto demanda algum tempo para a sua conclusão, muitas vezes, mais de dois anos, e concomitantemente à execução da avaliação, existe uma pressão iminente pela ampliação das políticas educacionais, não é algo banal manter um projeto avaliado *parado* enquanto uma avaliação de impacto é concluída. No relato abaixo o gerente do projeto comenta e exemplifica a coação pela ampliação de projetos educacionais, como aconteceu no projeto Reinventando o Ensino Médio – antes mesmo que alguma evidência sobre sua validade e resultados fosse concluída, o projeto já tinha um cronograma de alargamento definido.

> Então é meio complicado, porque a avaliação de impacto também você não pode fazer ela correndo, no mínimo, no mínimo você precisa de um a dois anos. E eu acho que na educação, se a gente for esperar um a dois anos para fazer uma avaliação de impacto como tem que ser feita, nós vamos perder dois anos de atividade. Igual, agora, o Reinventando o Ensino Médio, começou com 11 escolas, em 2013 vai para mais 122, em 2014, já vai para o estado todo. Como é que nós vamos esperar uma avaliação de impacto para poder universalizar? Então eu acho que na educação tem hora que é difícil fazer avaliação de impacto por este prisma que tem que universalizar e não pode demorar muito. Mas eu acho que ela é muito importante. (GERENTE DO PROJETO PAV)

O entrevistado SEE2 concorda com a observação do gerente na medida em que argumenta que as decisões na Secretaria de Educação são tomadas porque têm que ser tomadas e, não, porque existem evidências para tal. Para Shulha e Cousins (1997), a atividade política também estaria relacionada ao contexto de realização das avaliações, sendo que a tomada de decisões, alinhada às recomendações de uma avaliação, não depende somente de um único ator, mas de outros grupos de interesse e circunstâncias do próprio setor público. Nesse veio, Cousins e Leithwood (1986) ressaltam que valores e interesses de pessoas envolvidas nos programas mostraram-se concorrentes poderosos com os resultados das avaliações. O que condiz com a percepção do entrevistado SEE2.

> Aqui na Secretaria de Educação, as tomadas de decisões são feitas, não baseadas em dados técnicos; são feitas, porque têm que ser feitas. Você tem que tomar uma decisão em determinado momento, então você toma a decisão naquele momento na mesa: "vamos fazer isso, isso e isso". (EQUIPE SEE2)

Além disso, a pressa pela ampliação de programas antes mesmo que se tenha qualquer evidência sobre sua efetividade, além da tomada de decisões precipitada, não favorece o desenho aleatorizado das avaliações de impacto. A melhor maneira de evitar uma amostra enviesada ou não-representativa é selecionar uma amostra aleatória, definida como uma amostra em que a probabilidade de qualquer indivíduo da população ser selecionado é exatamente a mesma de qualquer outro indivíduo (KHANDKER; KOOLWAL; SAMAD, 2010). Para o gerente do projeto, não seria justo implantar o projeto em apenas algumas escolas, que seriam o grupo-tratamento, e deixar outras escolas sem o projeto que, por sua vez, comporiam o grupo-controle, devido a uma experiência para verificar o impacto. Nesse caso, questiona-se o dilema ético das avaliações de impacto em projetos sociais de escolher entre atores sociais, no caso alunos, que poderiam participar do projeto, como um experimento, e alunos que não deveriam participar do projeto. Como está exemplificado na fala do gerente do projeto, abaixo.

> A avaliação de impacto, ela é uma avaliação diferenciada, e eu acho que ajuda muito a gente a perceber se o projeto realmente está dando resultado ou não está. Por outro lado, como as políticas publicas da educação precisam ser políticas que abranja todo o universo das nossas escolas, eu acho que em alguns casos é difícil fazer avaliação de impacto na educação por isso. Vou te dar um exemplo. Como que eu posso, por exemplo, implantar um PIP, trabalhar o Programa Intervenção Pedagógica em "x" escolas e não em todas? Não tem como. Então, eu acho que a avaliação de impacto, ela é importante quando a gente tem projetos pontuais, como é o caso do PAV, mas, no geral, na educação a gente acha um pouco difícil não fazer projetos que não sejam universais. (GERENTE DO PROJETO PAV)

No entanto, em oposição ao problema ético das avaliações afirmado pelo gerente do projeto, os entrevistados do CEDEPLAR e do EpR ressaltam a importância de realizar avaliações com desenho aleatório *a priori*, dada a necessidade de encontrar conclusões mais confiáveis sobre o impacto dos projetos. Para Khandker, Koolwal e Samad (2010), é de se esperar que, escolhendo aleatoriamente grupos de tratamento e de controle, a probabilidade de determinadas características estarem presentes no grupo-tratamento seja a mesma de elas estarem presentes também no grupo-controle. Por outro lado, sem selecionar aleatoriamente, é possível fazer uma escolha de modo que as características observáveis sejam semelhantes, mas não é possível garantir que uma característica de difícil mensuração, como motivação, por exemplo, seja comparável nos dois grupos. De acordo com ambos entrevistados, o

desenho aleatório de políticas, desde o seu início, permite maior confiança nos resultados das avaliações de impacto e, assim, respostas mais embasadas e mais certas sobre a validade do projeto para um determinado grupo de beneficiários.

> Esse é o dilema. A pergunta é: se deveria continuar com a política pública que não dá resultado? Por outro lado tem outra dificuldade também o seguinte, porque como nada disso é feito de forma aleatória a gente tem que usar metodologias alternativas e tal. Claro, têm problemas essas metodologias, só que a gente propõe o aleatório e nunca consegue. Agora, por exemplo, a gente falou: "Vamos fazer aleatório", com a Secretaria de Educação, e, "Não, não, não, não". Então, depois não reclama, o programa vai demorar e não adianta, não dá para espremer os dados. Porque a inclusão de municípios é político. (COORDENAÇÃO CEDEPLAR)

> Os programas tinham problemas em relação a como serem avaliados, de origem. Para avaliação de impacto, por exemplo, é mais completa a avaliação se pudesse escolher entre os beneficiários de forma aleatória, por exemplo, e nunca nenhum desses projetos o público-alvo foi escolhido de forma aleatória que permitisse aos avaliadores de posse das ferramentas que eles tem, das metodologias de avaliação, fazer uma avaliação da política pública que pudesse ter recomendações generalizadas, que pudesse falar para a população como um todo e não apenas para aquele grupo especifico que foi avaliado. Seguindo o rigor acadêmico, os avaliadores estavam certos, por exemplo, em evitar fazer recomendações com base nos resultados, porque esses resultados não eram generalizados, e as recomendações seriam generalizadas para as políticas públicas. (TÉCNICO EPR3)

Além dos problemas inerentes à metodologia de avaliação de impacto, notam-se também dificuldades relacionadas a problemas estruturais próprios de projetos educacionais, de concepção e desenho do projeto. Na visão do entrevistado SEE2, a aferição dos resultados da avaliação é prejudicada, não por sua complexidade metodológica, mas devido à falta de foco dos projetos e `a ausência de objetivo e escopos de atuação claros, uma vez que escopos de atuação muito amplos tendem a abraçar muitas ações de natureza diferente que prejudicam a mensuração dos resultados podendo levar ao dispêndio de esforços e recursos. Caso a avaliação seja realizada em projetos com problemas de escopo muito amplos, as avaliações serão desperdiçadas, já que, *a priori*, os projetos não têm uma definição clara de sua estratégia de atuação.

> O que acontece muito nas avaliações de projeto é que, muitas vezes, o projeto, ele é desenhado e quer abraçar o mundo, então, ele quer resolver todos os problemas da educação com. Ele não quer só melhorar o desempenho do aluno, ele quer melhorar o desempenho, melhorar a autoestima, melhorar até a autoestima do professor, do pai, ele quer tudo. Aí quando eles vão desenhar, eles: "Olha, mas esse projeto está mal desenhado". Porque eles põem aqueles nobres da educação e aí ele vai olhar e: "Ah, mas o projeto não teve esse resultado!". Não teve porque não tinha como ter, porque não tinha uma ação que tivesse como chegar naquele resultado. Então, às vezes, também tem até um pouco de problema na concepção dos projetos – o projeto quer demais. (EQUIPE SEE2)

A ausência de diagnósticos que permitam identificar com clareza o problema a ser sanado pelo projeto e, consequentemente, a má definição de uma estratégia adequada de intervenção é também reprodução dos erros de uma cultura falha de planejamento, ainda comum no País, como argumenta Lobo (2001). O desdobramento do projeto em suas etapas, a melhor delimitação de seu objetivo e público-alvo, etapas essenciais a qualquer projeto público, nem sempre estão presentes na elaboração de um bom planejamento. E a inexistência de planejamento corrobora os problemas informacionais dos projetos avaliados, lacuna essa que compromete a factibilidade das avaliações. Para o entrevistado SEE1, a coleta de dados dos projetos educacionais, imprescindível às avaliações de impacto, era *arcaica* devido `a insuficiência de critérios de seleção e carência por sistemas confiáveis de dados. De acordo com o relato, até números básicos, como o quantitativo de alunos ou escolas não estavam sistematizados, como se explica nesta entrevista.

> Eles [os projetos avaliados] não estavam preparados. Porque, na realidade, em nenhum desses projetos não se foi pensando em uma avaliação de impacto no início, então até o controle mesmo de cada gerente, do que estava sendo atendido, do número de alunos atendidos, de escolas e tudo o mais, era um pouco... "manual", a palavra não é "manual", é um pouco arcaico no sentido assim, não era uma coisa padronizada. Não tinha bem definido qual que era a população-alvo, porque que aquela população era alvo. Faltava esse tipo de critério. As escolas, quais os critérios de escolher as escolas, os alunos, até o número de pessoas que estavam sendo atendidos, o número de alunos, o número de escolas, tudo isso não tinha um conhecimento certo, era pouco por alto. (EQUIPE SEE1)

As percepções dos entrevistados da Secretaria de Educação são corroboradas pelos pesquisadores, que afirmam que a indefinição de critérios para seleção dos beneficiados e a ausência de informações para a avaliação seriam fruto da maneira como os projetos são concebidos no governo. Na visão do entrevistado do CEDEPLAR, os programas governamentais são concebidos de acordo com ideias de um único ator, ideias estas que não são delineadas e especificadas, minimamente, antes da política pública ser lançada. Tal prática acarreta, posteriormente, as dificuldades quanto ao conteúdo das avaliações. Na analogia do entrevistado, no momento de se fazer a avaliação dos projetos inventados, o trabalho árduo do avaliador é tentar separar grãos de areia misturados desde o início.

> Avaliar uma coisa que já está pronta e a gente nem sabe como é que começou direito. Quer dizer, o gestor tenta passar o máximo de informação possível, a gente tenta extrair o máximo possível, mas não é igual a um processo que a gente tivesse iniciado antes, concebido política, definido na concepção da política quais as bases de dados utilizados, definir se aquela política é boa ou se não é boa, se existe alguma experiência internacional, como é que foi feito, que tipo de base de dados que precisa, qual que o sistema de dados... Então, assim, é sempre muito mais difícil, é

> muito esforço, e aí acaba ficando mais caro, você tem mais horas de trabalho. Aí não sei se é culpa do governo. Acho que é uma coisa geral do Brasil. Eu fico imaginando, deve ter "trocentas" politicazinha por aí, programas criados assim. Têm uns dez programas assim. Então está lá. O Estado, como o negócio já começou, ele adota aquilo sem saber. O cara vai lá com o negócio para enfiar na cabeça do cara, aí a gente chega para avaliar, o negócio foi criado da cabeça do cara, você não sabe o que que o cara criou, qual o critério. Fica todo mundo com grande dificuldade de definir que critério que foram utilizados para escolher as escolas ou os indivíduos. É importante isso para definir a avaliação. A gente tem essa dificuldade. Mas quando a gente consegue fazer o máximo para separar, a coisa está misturada como se fosse areia misturado com outra coisa, a gente fica tentando separar os grãozinhos, mas já foi misturado desde o início. Então, assim, isso tudo a gente deixa muito claro nas avaliações que isso tudo pode atrapalhar a avaliação. (COORDENAÇÃO CEDEPLAR)

Todos os problemas relativos aos projetos e ao sistema de dados da Secretaria de Educação, no entanto, não eram vistos como empecilhos à realização das avaliações, de acordo com os dirigentes da SEE. Na fala de um dos secretários de educação, *a estratégia de avaliação de impacto teria que se adaptar `a estratégia do projeto e, não, o contrário*, como relata o entrevistado da equipe da Superintendência de Informações Educacionais.

> Ele até criticava, porque o pessoal [do EpR] falava com ele [Secretário de Educação à época]: "Ah, a gente tem que fazer avaliação de impacto, mas a gente tem que fazer escolha aleatória". Já ouvi ele falando inclusive: "Eu não vou adaptar minha estratégia para uma avaliação de impacto. A avaliação de impacto que tem que adaptar a estratégia da secretaria". Eu já ouvi ele falando isso. E, de certa forma, ele não deixa de ter razão, né? Ele não deixa de ter razão. (EQUIPE SEE2, grifo da autora da dissertação)

E, assim, devido à tradição da concepção imediata dos projetos no governo, sem o mínimo diagnóstico, somada à ausência de sistemas de informações e dados do projeto, as avaliações tornam-se estudos frágeis, com conclusões que pouco podem transformar a realidade. Para Cousins e Leithwood (1986), para que as avaliações sejam utilizadas, é essencial que os resultados não sejam conflituosos e que os usuários tenham confiança nas informações geradas pelos estudos. A falta de credibilidade das conclusões do relatório afeta negativamente o uso, como adianta o entrevistado. Logo, como comentado a seguir, os usuários apresentam postura indiferente aos resultados da avaliação caso esses não sejam seguros: *se fosse positivo estava bem, negativo, porque não foi bem.*

> Não foi bem desenhado não porque as pessoas que fizeram não sabiam, mas é porque realmente era difícil fazer porque não tinha critério de escolha dos alunos, não tinha informação suficiente em nível socioeconômico com os alunos. Foi difícil. Como foi difícil fazer um desenho, aí acaba que o resultado assim: se fosse positivo estava bem, negativo, porque não foi bem. Então, não foi utilizado. (EQUIPE SEE1)

A insegurança nos resultados, do ponto de vista do pesquisador do CEDEPLAR, é, de fato, um desafio. Interromper um projeto somente devido à avaliação de impacto é uma atitude muito arriscada: primeiro, porque o pesquisador nunca terá certeza absoluta quanto ao resultado, devido às dificuldades metodológicas com que se depara ao longo do caminho e segundo, dado que a responsabilidade de realizar decisões severas quanto às políticas de um governo tem implicações políticas densas. Para elucidar a ideia, o entrevistado do CEDEPLAR emprega um exemplo hipotético para o projeto Bolsa Família, sendo sua percepção compartilhada por entrevistados da coordenação do EpR.

> Tem sempre essa dificuldade, porque a gente não tem certeza absoluta, tem insegurança nos resultados. Esse é um desafio. Quer dizer, o dado que você tem no resultado que me mostra que não houve impacto... Imagina isso acontecendo com o Bolsa Família, por exemplo, e aí não deu impacto, o governo vai ignorar, não adianta, porque aquilo ali politicamente... Ele vai sempre falar: "Não, mas sua avaliação podia ser melhor por isso, por isso...". O cara vai sempre se prendendo por isso, que talvez até seja verdadeiro, porque você tem dificuldades metodológicas, mas aí o troço já expandiu, já virou um monstro enorme e tal. (COORDENAÇÃO CEDEPLAR)

> E o discurso acadêmico de uma avaliação de impacto, ele não está bem casado com o discurso político. Por quê? Porque tem uma dissociação, ou seja, isso aqui é muito técnico, ele tem muito 'se, se, se, se...' tem muita hipótese. Política é um negócio que ele vai sentar e gritar para os quatro ventos. Mas, se usou, poderia ter usado mais? Sim. O político poderia ter entrado mais um pouco nesse aspecto e levantado um pouco mais a bandeira. Mas não é fácil. (COORDENAÇÃO EPR1)

> Então há uma certa tensão entre a capacidade de instrumentos técnicos, como monitoramento e avaliação produzir resultados efetivos, ou seja, orientar a alocação de recursos. (COORDENAÇÃO EPR2, grifo da autora da dissertação)

E, assim, em meio aos inúmeros problemas enfrentados na fase de implementação e execução das avaliações, sejam estes de envolvimento dos atores, dificuldades metodológicas, ausência de dados, falta de conhecimentos dos gestores, características contextuais, entre outros, que danificam etapas posteriores das avaliações. A apresentação dos resultados da avaliação do PAV e o processamento de informações dela recorrente são foco da seção seguinte. É importante observar que não será exposta nenhuma referência ao relatório de avaliação do PAV, devido à cláusula de sigilo presente no contrato das avaliações educacionais.

10.3 Apresentação dos resultados e processamento de informações

Leviton e Hughes (1981) enfatizam a importância de submeter os resultados de uma avaliação em tempo de se tomar as decisões necessárias, principalmente para o uso instrumental; em

outras palavras, os autores citam a importância da tempestividade para o se uso. Quanto a essa característica, os autores citam, não somente o tempo de espera para a apreciação dos resultados, mas também a interferência da fase do ciclo político em que as avaliações são recebidas. E, de fato, diferentes entrevistados comentaram a relação entre a ocasião na qual os resultados foram entregues e a utilização das avaliações. Na visão da Secretaria de Educação, é próprio da natureza das avaliações de impacto ser mais delongada, o que nem sempre coincide com o ciclo político do governo, como explica o entrevistado da SEE e os coordenadores do EpR.

> Um dos problemas que eu acho da avaliação de impacto é que o tempo da avaliação de impacto não coincide muito com o tempo de governo, porque para a avaliação de impacto você tem que ter no mínimo dois anos e, no governo, tudo é muito rápido as coisas. Ás vezes, não há o tempo necessário, porque você tem que fazer, tem que mudar a política. Tem um *gap* entre tempo necessário para fazer uma boa avaliação e tempo que precisa para política pública. Acho que essa área é uma dificuldade das avaliações. (EQUIPE SEE1)

> E isso acontece, os resultados da avaliação, eles geralmente não casam com o tempo do governo, então eles não vão estar prontos quando precisa, demora um pouquinho, quando eles veem já tem mudança de governo. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

> O ciclo político, ele é curto. Ele é, no máximo, cinco anos em alguns países, mas no nosso país são 4 anos, sendo que ele é mais curto ainda porque já no último ano o grupo que detém o poder já está preocupado com a nova eleição, e é natural da democracia, eles precisam se ocupar com o seu capital político – o capital político dele é dado pelas apostas que eles fizeram há "x" anos atrás. Então, mesmo que as avaliações de impacto tivessem mostrando resultados ruins, mas se a aposta do político, que ele fez em determinado programa, isso não vai voltar atrás porque é a imagem dele, é a reputação dele. (COORDENAÇÃO EpR2)

A avaliação do PAV não fugiu à regra das avaliações de impacto. De acordo com relatos, a avaliação, que levou dois anos aproximadamente para ser concluída, apresentou atrasos consideráveis. O retardamento na conclusão da pesquisa, somado à falta de contato entre avaliadores e avaliados ao longo de seu desenvolvimento, chegou a suscitar dúvidas quanto `a concretização da avaliação por parte do gerente do projeto, como comenta o entrevistado do EpR no fragmento de entrevista abaixo. De acordo com Bamberger, Rugh e Mabry (2011), os usuários de uma avaliação devem manter-se informados durante todas as fases de um estudo avaliatório, sendo avisados a cada descoberta preliminar feitas pelos avaliadores. Contudo, no caso dos projetos educacionais, isso não aconteceu. Quando, de fato, as avaliações foram concluídas, o entrevistado do EpR relata que muitos dos projetos da Secretaria de Educação já apresentavam mudanças antes mesmo de ter as evidências da avaliação.

> Todos os produtos atrasaram muito. Eu me lembro quando, conversando com o gerente do PAV sobre o resultado da avaliação de um determinado momento, até falou assim: "Ah, achei que tinham cancelado essa proposta, essa iniciativa, porque eu não tive mais nenhum contato a equipe da UFMG nesse um ano", sendo que dois anos que já tinha se passado desde quando foram consultados. A informação quando chegou, dois anos depois, já era defasada, informação velha, pois muitos projetos já tinham passado por mudanças significativas. (TÉCNICO EPR3)

Tal informação coincide com a percepção dos técnicos da Secretaria de Educação, comentada na seção anterior, quanto *trade off* entre a extensão para tempo de execução da pesquisa e a necessidade urgente de expansão de projetos educacionais, já que não é algo banal manter um projeto avaliado *parado* enquanto uma avaliação de impacto é concluída. Além disso, o momento em que os resultados chegaram aconteceu na transição entre dois governo, período marcado pela rotatividade de pessoas, em especial dos *stakeholders* envolvidos na avaliação, tanto no EpR, quanto na Secretaria de Educação. Os comentários dos membros da Superintendência de Informações Educacionais apontam que a ocasião foi uma *confusão danada*.

> Até porque eu acho que depois teve algumas revisões e acho que demorou um pouco. Mas depois a gente foi perdendo um pouco de contato também, na época o Estado para Resultados estava acabando. Foi uma confusão danada. Isso foi o que eu senti, mas não sei se teve outras reuniões só com os gerentes, isso eu não posso falar porque foi justamente nesse processo em que eu estava saindo da superintendência. E também lá no escritório, as pessoas que estavam com contato direto saíram, então teve essa falta de diálogo, não sei bem o que que é isso. (EQUIPE SEE2)

> Primeiro, que, mudou lá no escritório e algumas pessoas saíram. As pessoas que estavam envolvidas no começo saíram, entraram outras, então, perdeu um pouco o contato e não teve, depois, uma discussão mais extensa. Então, eu achei que perdeu muito. Na verdade, não foi aproveitado. (EQUIPE SEE1)

Diferentes atores fazem convergir as percepções quanto `a ausência de uma apresentação formal dos resultados da avaliação, que propiciasse a discussão mais profunda sobre os projetos educacionais avaliados, dentre eles o PAV. A disseminação dos resultados da avaliação se manteve ao nível gerencial da Secretaria e EpR, o que poderia ter restringido o processamento de informações, como no fragmento de entrevista abaixo.

> Aí fizeram as avaliações e não teve uma divulgação aqui dos resultados, divulgação ampla, eu falo assim, chamar todos os atores. É. "Olha, aqui foi o resultado do PAV", não, a divulgação ficou mais no nível gerencial; passaram para o secretário, o secretário adjunto, na época, e tal. Eu acho que faltou isso, tinha que reunir as pessoas da secretaria. (EQUIPE SEE2)

Como apresentado nos padrões do *Joint Commitee for Educational Evaluation* (1994), a informação coletada pela avaliação deveria ser selecionada para responder amplamente as questões pertinentes sobre o programa e ser responsiva às necessidades e interesses dos clientes e outros *stakeholders* envolvidos. A ressalva existente no modelo de avaliações coordenada pelo governo mineiro, contudo, coloca o EpR como cliente das avaliações, ao contrários dos gerentes e técnicos do projetos, pessoas que, de fato, conhecem e executam o projeto. Essa classificação do EpR como cliente da avaliação cria uma distância entre pesquisadores e avaliados, inclusive, no momento de divulgação dos resultados. Como relatam os entrevistados do EpR e do CEDEPLAR nos trechos abaixo, via de regra, as apresentações dos resultados eram feitas para o EpR, que também cuidava da leitura final do relatório e, somente depois, o próprio EpR apresentava para a SEE, de acordo com a conveniência.

> Via de regra, a apresentação era para o Estado para Resultados, inclusive, a gente lia o relatório primeiro, então, a gente revisava o relatório e pedia correções, esclarecimentos do que a gente considerava insuficiente. Erros triviais de digitação ou de português porque não foram bastante cuidadosos com o relatório final, até um maior detalhamento do por que algumas recomendações estavam sendo dadas ou no que elas eram fundamentadas. Depois teve a apresentação, que aí era constituída mesmo no *Power Point*, o Estado para Resultados apresentava para a gerência do programa na Secretaria de Educação. Acho que foi basicamente só isso que ocorreu para disseminação. (TÉCNICO EPR3)

> Quase nada. Eu lembro assim mais do Poupança Jovem. Esses três menores [PAV, PEP e Escola Integral] entregamos relatórios, um técnico do EpR falou que ia marcar e não marcou. Mas não é culpa dele, houve as tentativas e não deu. (COORDENAÇÃO CEDEPLAR)

Para Bamberger, Rugh e Mabry (2011), a chance de reagir, por parte dos usuários de uma avaliação, quando apresentadas as conclusões e recomendações, também constitui parte essencial do trabalho. No entanto, como afirmado pelos entrevistados da SEE, não houve uma discussão que motivasse o desdobramento das conclusões da avaliação.

> Os gerentes leram e participaram. Mas igual eu estou falando, leram e participaram, mas você sabe quando lê um trabalho, aquele negócio e tal e não há nenhum desdobramento daquilo. Assim, foi uma apresentação, não teve uma discussão, pelo menos que eu tenha participado. (EQUIPE SEE1)

> O que faltou mesmo, eu acho, que foi a divulgação e internalizar mesmo esse conhecimento aqui, tentar explicar a importância, colocar para as pessoas: "Olha, vamos ficar mais atento nesses resultados na hora de a gente tomar decisão". (EQUIPE SEE2)

Ademais, os atores comentaram a dificuldade de acesso aos relatórios finais das avaliações. Mesmo que o JCSEE (1994) aponte que os formalmente envolvidos na avaliação deveriam garantir que todas as descobertas derivadas da avaliação se tornassem acessíveis às pessoas afetadas pela avaliação e a quaisquer outros com expressos direitos legais de receber os resultados, a acessibilidade dos relatórios, assim como a apresentação, foi restrita. Como exemplo, membros da Superintendência de Informações Educacionais, que trabalharam diretamente no fornecimento de dados e contato com EpR e CEDEPLAR, confessam que custaram a receber os relatórios e que, para ter acesso aos informes, tiveram que solicitar.

> Para começar, eu custei a receber os relatórios. Eu tive que pedir para ver os relatórios. Eu pedi para o pessoal que saiu do escritório e virou... Não! Como é que é? Era Estado para Resultados e virou escritório, então eu pedi para o pessoal... Nessa mudança é que perdeu um pouco, sabe. (EQUIPE SEE1)

> Mas fato é que não teve uma apresentação nem uma divulgação até para os setores externos. Se alguém quiser ter acesso a essas avaliações é até complicado. Até pouco tempo atrás poucas pessoas aqui da Secretaria tinham esses arquivos. Por exemplo, eu tenho porque eu pedi, eles não passaram para a gente; "Olha, a gente está entregando as avaliações...". (EQUIPE SEE2)

Somada à dificuldade de acesso, acomete a avaliação a ausência de incentivos para facilitar o entendimento dos atores quanto aos relatórios de avaliação, que trazem no corpo do texto fórmulas e termos complexos inerentes às metodologias que mensuram o impacto de políticas. Como corretamente assinalado por Leviton e Hughes (1981), a utilização pode ser prejudicada se os incentivos para ler os relatórios de avaliação são baixos. Como exemplo, os autores citam os jargões e vocabulários específicos utilizados pelos avaliadores como um dificultador do entendimento dos resultados das avaliações pelos usuários, uma vez que, para a maioria dos gestores e políticos, a avaliação ainda constitui uma ferramenta incomum. O entrevistado da equipe da SEE comprova o fato, quando comenta que se trata de um relatório complicado, de difícil entendimento.

> O relatório é muito complicado, mas você consegue ler as conclusões, mas aquela parte de econometria, eu, por exemplo, não entendo quase nada. A gente lembra de algumas coisas que a gente estudou, mas é bem superficial, tem uma ideia por onde está passando aquilo. Quem quiser ler pode pular aquela parte e ler introdução e conclusão, pelo menos você vai saber: "Deu impacto. Deu, mas não deu muito...". Mas é um relatório complicado, é difícil para as pessoas. (EQUIPE SEE2)

Por fim, em síntese, o gerente do projeto recupera o processo percorrido pelas avaliações na Secretaria de Educação em duas etapas bem delimitadas, como narrado no relato abaixo. Nas

suas palavras, primeiro, o gerente foi informado pela alta direção da Secretaria que haveria uma avaliação; segundo, já na etapa de conclusão da avaliação, recebeu o relatório final de avaliação de impacto para análise. A simplicidade com que o entrevistado descreve o processo de avaliação é indiscutível, o que pode comprovar a ausência de interação entre avaliadores e gestores dos projetos avaliados comentada por outros *stakeholders*. Quanto ao relatório final de avaliação, o entrevistado confessa que lhe falta memória se ainda tem o documento.

> A alta direção da Secretaria da época chamou a gente e disse que alguns projetos estruturadores teriam uma avaliação de impacto e foram escolhidos o PAV e parece-me que o PROETI. Não tenho certeza não, mas eu acho que foram esses dois. Esse foi o primeiro contato que a gente teve. Depois disso, nós recebemos um documento elaborado pela consultoria que fez a avaliação de impacto para a gente analisar, pegou o documento para a gente ler e analisar. Recebi, li, analisei, tivemos uma reunião em que a gente levantou algumas questões, não me lembro mais quais são, algumas questões que a gente não estava entendendo muito bem. Basicamente foi assim, foi uma decisão da secretaria de avaliar os projetos estruturadores, na época o PAV era um projeto estruturador, depois a gente recebeu os documentos, analisamos. Agora, o documento final eu nem sei se eu tenho ele ali. Não tenho muita certeza não. (GERENTE DO PROJETO PAV)

A lacuna de discussão dos relatórios de avaliação, somada aos demais problemas encontrados ao longo da avaliação de impacto dos projetos, levou, na percepção do entrevistado SEE1, a certa indiferença dos gestores dos projetos quanto ao processo. Nota-se, portanto, apatia da Secretaria de Educação: se os resultados fossem positivos, os gerentes os recebiam bem, mas não havia mudança associada ou até *eu já sabia*; se os resultados fossem negativos, empregava-se a desculpa de que o desenho do projeto ou os dados utilizados não eram adequados o suficiente e, por isso, a avaliação não tinha credibilidade. Tal percepção é comentada no trecho de entrevista abaixo.

> A diferença foi assim, foi positivo, beleza, não mudou nada e tal. Negativo foi assim: "Ah, deu negativo porque o desenho não foi bem feito". Então, assim, não gerou discussão e as pessoas receberam o gerente dessa forma: "Foi negativo porque o desenho não estava bom, então por isso que não mostrou resultado positivo". E o outro o resultado foi positivo: "Ah, beleza. Eu já sabia". Mais ou menos isso que eu senti. (EQUIPE SEE1, grifo da autora da dissertação)

Na percepção do entrevistado, os relatórios de avaliação em nada influenciaram a tomada de decisões na secretaria: *eu vi isso na prática, esses relatórios não influenciaram nenhuma tomada de decisão aqui dentro, fato*. Essa visão será mais bem explorada na próxima seção deste trabalho, seção esta que trata da implantação das recomendações decorrentes da avaliação do PAV.

10.4 Utilização da avaliação do ponto de vista dos *stakeholders*

Embora não seja possível divulgar os resultados reais da avaliação do PAV neste trabalho, em respeito à cláusula de confiabilidade presente no contrato entre governo e CEDEPLAR, é possível observar diferentes percepções quanto à utilização das avaliações educacionais. A maior parte dos atores demonstrou certa frustração quanto ao processo, ressaltando a influência política em detrimento do técnico. Como pontos positivos, os entrevistados concordam com o aprendizado e a internalização dos conhecimentos sobre o tema de avaliações, tanto por parte dos avaliadores, quanto dos avaliados, e a melhor organização dos dados administrativos dos projetos após a ínfima experiência de conseguir os dados necessários à execução dos estudos.

Na visão do entrevistado do CEDEPLAR, a mobilização que incentivava a realização das avaliações no governo mineiro, experimentada no início de 2007, se perdeu. Um dos motivos para essa *desmobilização*, na percepção do entrevistado, diz respeito à ascensão mundial do tema de avaliações, que impacto que existia no início das discussões e que, atualmente, não apresentaria a mesma importância no governo. Os incentivos criados à época, somados à presença do Banco Mundial, além das várias avaliações de impacto espalhadas em diferentes governos no mundo – como a avaliação do programa Progresssa[12] no México –, foram passageiros e não estariam, assim, mais na pauta de prioridades do governo. A impressão do entrevistado é de que a preocupação com o tema de avaliações não passaria de uma tendência, altamente mutável, como a *moda*.

> Se perdeu. Gera-se uma cultura, está começando a produzir a cultura da avaliação, mesmo que não fosse avaliação de impacto, mas acompanhamento crítico da política, meio que isso vai se perdendo também se você não intensifica. Depois, isso foi desmontado. Desmontando no sentido de desmobilizado. Você tinha uma mobilização no início para a avaliação técnica, todo mundo interessado, o Banco Mundial junto incentivando. Aquilo que eu te falei; tinha uma mobilização e aí se perdeu. Claro também que a potência internacional era esse, tinham muitos programas sendo avaliado no mundo, o Progressa no México e tal. Foi uma expansão. Agora essa coisa passou um pouco, essa é a minha sensação. É uma certa moda, com certeza. A minha sensação é que havia mais, muito interesse nas avaliações. (COORDENAÇÃO CEDEPLAR)

[12] O Progressa constitui um dos programas sociais do governo federal do México que foi avaliado de acordo com os padrões de avaliação de impacto aleatório e que, portanto, apresentou os resultados mais contundentes.

Nota-se, assim, certo sentimento de frustração quanto ao aproveitamento e utilização das avaliações educacionais, na percepção do pesquisador. A mudança de interesses e de prioridades do governo quanto às avaliações, marcada pela diferença de mobilização existente no início das discussões e no período de fechamento das avaliações, é cunhada pelos autores Leviton e Hughes (1981) como *advocacy*, aqui entendido como o apoio alcançado frente ao contexto político em que as avaliações são conduzidas. O apoio político e compromisso de um ou mais indivíduos-chave com a avaliação seriam fundamentais para a utilização dos seus resultados (LEVITON; HUGHES, 1981). Nesse veio, a tomada de decisões, alinhada às recomendações de uma avaliação, não dependeria somente do apoio e decisão de um único ator, mas de outros grupos de interesse e circunstâncias do próprio setor público, como também revelam Shulha e Cousins (1997). A extinção desde apoio, comentada pelos autores, em contrapartida, levaria as avaliações ao fracasso. E o pesquisador completa, uma de as razões das avaliações em Minas caírem em desuso está relacionada aos aspectos políticos da tomada de decisões frente à função técnica da avaliação. Como exemplificado no trecho de entrevista abaixo, a parte técnica de uma avaliação perde força frente `a pressão política de atores influentes no governo, percepção também corroborada pelo entrevistado do EpR.

> Mas o aspecto política, ele sempre pesa muito para definir quem vai receber e quem não vai receber, o próprio desenho. Se tem uma pessoa forte dentro do governo vai falar "Não, eu vou manter esse programa porque eu que inventei". Isso prevalece. Então tem essas coisas. Por isso que a parte técnica acaba perdendo força. (COORDENAÇÃO CEDEPLAR, grifo da autora da dissertação)

> Ninguém que tinha autonomia para deixá-las fora das gavetas queria que elas ficassem fora das gavetas, eu acho. Mas a minha impressão é de que ninguém estava disposto a comprar essa briga na época ou permitir que outras pessoas comprassem essa briga também usando as informações que a avaliação disponibilizava. Como a posição, por exemplo, teria sido um tiro pela culatra: teriam tido a intenção de avaliar as políticas para justificar excelentes ações e estariam municiando a oposição para bombardear esses projetos. Acho que foi tendo isso sim, uma percepção clara de que talvez o interesse maior fosse de natureza política e não técnica, e o lado político disso não estaria disposto a arcar com os custos do que uma decisão técnica que orientasse. (TÉCNICO EPR3, grifo da autora da dissertação)

No relato abaixo, no entanto, o entrevistado comenta que, embora as avaliações não tenham sido utilizadas, a experiência gerou aprendizado na confecção de avaliações, principalmente no bom emprego de secundários para a pesquisa de projetos sociais.

> A coisa começou muito bem com essa ideia de a gente primeiro tentar construir propostas de políticas e depois fazer a avaliação. Um pouco frustrante, porque nada foi usado, mas a gente ainda sim insistiu, assim, como fazer as avaliações mesmo de programas que a gente não tinha proposto, mas o que tem de bastante interessante aí

> é exatamente o desenvolvimento de metodologias para avaliação especialmente do uso de dados secundários. Porque, no caso do Poupança Jovem, a gente coletou os dados, mas, mesmo assim, a gente precisou usar dados secundários: dados do senso escolar, usar dados da Secretaria de Educação. E para os outros programas a gente não coletou dado nenhum, eram só dados secundários. Então, essa experiência foi muito positiva, eu acho. Só acho que ela foi pouco aproveitada. O CEDEPLAR, por exemplo, acho que ganhou muito. Ele criou um volume de informações grande que para ele mesmo é bom. É o velho dilema de quem mais ganha é o consultor, porque ele ganha duas vezes. Eu acho que se criou uma massa crítica importante nesse sentido. Essa informação vale um dinheiro. (COORDENAÇÃO CEDEPLAR, grifo da autora da dissertação)

Para o entrevistado do EpR, não é possível encontrar nenhum exemplo do uso das recomendações no PAV, já que as mudanças apontadas no relatório já estariam em curso antes mesmo do fim da avaliação, como afirmado nos fragmentos de entrevista abaixo. Na sua visão, as avaliações ficaram perdidas e caíram no esquecimento dos gerentes dos projetos avaliados. Para o entrevistado, nenhum gerente nunca se questionou a melhorar os projetos a partir dos dados das avaliações, porque eles nem lembram que elas existiram e que poderiam ser fonte de reflexão. E a falta de memória seria, até certo ponto, natural, visto que os próprios coordenadores dos projetos e as pessoas da área não foram incluídos no processo de avaliação, como se comenta nesta entrevista.

> Os resultados da avaliação foram intempestivos, eles demoraram demais. Acho que eu não consigo levantar um ponto que seja para poder falar que isso aconteceu por causa da avaliação, se a melhoria foi causada pela avaliação de impacto. Eu acho que não. Nenhum. Na verdade, a recomendação que tinha lá era uma coisa que já vinha sendo feita. A do PAV, que era de ampliar o programa também já estava sendo feita. Ele tinha sido pensado para o grande norte de Minas Gerais e foi ampliado para todas as regiões. (TÉCNICO EPR3, grifo da autora da dissertação)

> Ninguém nunca comentou, falou: "Ah, vamos melhorar o projeto porque aquela avaliação..." ou "Você lembra da avaliação de impacto? Mostrou isso...". Ninguém nunca toca nesse assunto. Por exemplo, quando você vai pensar numa reformulação desses projetos ninguém fala na avaliação de impacto, ninguém nem lembra que teve; os gerentes, o pessoal da área não lembra, porque também eles não acompanharam tanto, e acaba que a decisão de reformular, das estratégias, as ações desses projetos ficam na mão. Só que como eles não participaram nem da construção nem do andamento das avaliações e nem tampouco receberam informações, não chegou para eles qual foi o resultado, o relatório praticamente não foi entregue a eles, eu acho... Acho, não. Eles não utilizaram. Eu vi isso na prática. (EQUIPE SEE2, grifo da autora da dissertação)

Na visão do entrevistado do EpR, portanto, comprova-se o não uso da avaliação do PAV, bem como das demais avaliações educacionais que ocorreram no período. Apropriando-se da classificação feita por Alkin e Coyle (1988), observa-se um tipo de não uso intencional, que acontece no caso, por exemplo, de desconhecimento dos usuários ou pouco aproveitamento dos relatórios. Assim, a mesma sensação de frustração, de não aproveitamento das avaliações,

experimentada pelos pesquisadores do CEDEPLAR, não teria acometido os técnicos da Secretaria de Educação, uma vez que não havia expectativas por parte deles, não havia uma intenção preexistente de utilizar os dados, como comentado no primeiro trecho de entrevista. Para o entrevistado da coordenação do EpR, o nível de parceria e interação com a Secretaria de Educação era baixo, o que prejudicou a criação de capacidade de avaliação, como comentado no segundo trecho de entrevista.

> Por parte da Secretaria de Educação acho que não teve uma frustração tão grande porque eles não tinham grandes expectativas, eles deixaram acontecer e falaram assim: "Ok. Então pode fazer. Já não está na minha alçada decidir se vai fazer ou não, que direção vai, pode fazer, mas eu também não vou participar muito dessa confecção dessa avaliação não". E, de fato, eles não acompanhavam de perto, não participavam de perto. Eles poderiam ter tido outros momentos em que haveria essa retroalimentação intermediária. Por exemplo, "Ok. Os registros administrativos estão ruins? Vamos melhorar os registros administrativos para o ano que vem que aí para a próxima coorte a gente vai ter informação mais qualificada". Não fizeram isso, ou, não fizemos isso. Então na Secretaria de Educação acho que não teve frustração grande porque não tinham expectativa. (TÉCNICO EPR3, grifo da autora da dissertação)

> Então, nesse sentido, não se desenvolveu tanto a capacidade como foi o intuito da FJP e, principalmente porque a interação era muito ruim com a educação. Não que era muito ruim, mas o nível de abertura e de parceria com a educação era muito baixo com os dirigentes da secretaria de educação porque eles não estavam abertos para a gente ir. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

A percepção dos *stakeholders* da Secretaria de Educação quanto à utilidade da avaliação do PAV é controversa. Enquanto os entrevistados da Superintendência de Informações Educacionais emitem a opinião de que os resultados não foram utilizados; o gerente do projeto emite a versão de uma avaliação útil e satisfatória. Os entrevistados das equipes SEE1 e 2 comentam que não se discutiu o redesenho do programa devido `a avaliação, não só na avaliação do PAV, mas também nos demais estudos educacionais. Embora o entrevistado SEE2 tenha dito que os resultados foram positivos, cabe dizer que o projeto PAV não é mais um projeto estruturador, como também aconteceu com outros projetos avaliados do grupo. Assim, na percepção das equipes de SEE, os resultados não foram utilizados.

> Não vi muita utilização dos resultados. Foi positivo, ótimo e tal. O que foi negativo não se discutiu o redesenho do programa. (EQUIPE SEE1, grifo da autora da dissertação)

> Agora, o que eu vi em relação a isso é que não se utilizou dessas avaliações para tomadas de decisões, principalmente agora, futuras. Por exemplo, o PAV, que é um projeto, na época eu lembro que li a avaliação, foi o que apresentou um resultado mais significativo, e ele era um projeto estruturador. Ele deixou de ser estruturador.

> Mas, assim, eu não vi nenhuma mudança nesses outros dois projetos que foram o PEP e Escola de Tempo Integral em função do resultado da avaliação de impacto. O PEP continuou com a mesma lógica, não mudou nada. O Tempo Integral teve uma pequena mudança, mas que, no meu entendimento, não mudou praticamente nada. Simplesmente por decisão política. Assim, reformulou o projeto, mas na minha visão praticamente o projeto continua o mesmo acho que até na visão do gerente também. (EQUIPE SEE2, grifo da autora da dissertação)

Em oposição aos colegas, na visão do gerente do projeto, a avaliação foi positiva, principalmente porque mostrou que o projeto estava *cumprindo seu papel*. Assim, a avaliação teria certificado a expectativa inicial de que os resultados fossem positivos, perspectiva inicial do gerente, como comentado na primeira seção deste capítulo. Quanto à aplicação das recomendações no projeto, o gerente diz não se lembrar, teria que recorrer ao relatório para fazer a associação.

> Então eu achei que foi muito bem feita, gostei dos estudos, das análises que eles fizeram, pela avaliação que ele nos mostrou o PAV estava cumprindo o papel; os resultados foram positivos. Tinha sugestões, recomendações, o que devia melhorar no projeto. Mas eu não estou lembrando, não estou mesmo. Mas, assim, tem um tempo já maior e eu não estou me lembrando especificamente para te dar exemplo do que a gente melhorou do projeto, o que a gente não melhorou a partir dessa avaliação. Eu tinha que pegar o documento e ver de novo. Não lembro. (GERENTE DO PROJETO PAV)

Nota-se, assim, que a simpatia do gerente quanto à avaliação se refere mais à natureza positiva dos resultados que à possibilidade de poder refletir sobre o projeto a partir de dados técnicos e seguros. A insegurança e o desuso da avaliação podem ser fruto também da confiança depositada nas avaliações, como revelam Leviton e Hughes (1981), e comenta o entrevistado da SEE. De acordo com a percepção do entrevistado, se a avaliação tivesse sido demandada pela Secretaria, ao contrário de ser contratada por outra instituição de governo, como o Estado para Resultados, o trabalho poderia ter um fim diferente do acontecido. Como clientes verdadeiros da avaliação, seria mais fácil sensibilizar os gestores sobre a relevância das informações geradas por uma avaliação e torná-los menos resistentes quanto ao processo de trabalho. Assim, na hipótese do entrevistado, se a avaliação tivesse *nascido* na SEE, teria mais credibilidade por parte dos gestores dos projetos e, em consequência, melhor utilização.

> Eu acho que a partir do momento que você não envolve principalmente a gerência do projeto e as pessoas responsáveis pela tomada de decisão, acaba em desuso... Por isso que eu acho que uma avaliação de impacto com iniciativa da Secretaria de Educação você vai ter um pouco mais de crédito porque vai nascer aqui, as pessoas vão participar desde o início. Agora, quando chega de fora... Eles sentem muito mais uma avaliação externa medindo o trabalho deles. A partir do momento que você faz uma avaliação aqui, nascendo aqui você tenta explicar para eles desde o início: "Olha, a gente não está querendo saber se o seu trabalho é ruim ou bom, isso aí é

> justamente para a gente melhorar as políticas públicas...". Então há um processo de convencimento maior do que quando você já chega com o relatório pronto. Às vezes, a pessoa fica até com raiva: "Pô, mas o projeto não deu resultado, não sei o que. Que relatório é esse?". As pessoas ficam muito na defensiva também, então tem ás vezes um receio: "Olha, está fazendo avaliação...", já ficam com o pé atrás. (EQUIPE SEE2, grifo da autora da dissertação)

Embora a forma como as avaliações educacionais se encerraram tenha gerado frustações em alguns atores envolvidos no processo, como é possível perceber pelos relatos aqui expostos, eles reconhecem pontos positivos quanto à experiência. A primeira observação diz respeito à maneira de receber as avaliações por parte da Secretaria de Educação, como relatam os entrevistados do EpR, SEE e coordenação do CEDEPLAR. Para o técnico do EpR, atualmente, as pessoas da SEE convivem muito melhor com o *termo avaliação*. O habito foi fruto da coexistência de diferentes praticas de avaliação inseridas na Secretaria como o SIMAVE. Na visão da coordenação do CEDEPLAR, mesmo que, de início, a instituição tenha apresentado certa resistência ao trabalho, alguma pessoas começaram a demonstrar alterações comportamentais e cognitivas, como resultado da participação no processo de avaliação, ainda que tal participação tenha sido reduzida. Por fim, o entrevistado da SEE comunica a realização de uma nova avaliação de impacto encomendada por eles, demanda esta que talvez não fosse possível caso as avaliações educacionais não tivessem ocorrido, como atestam os entrevistados, respectivamente.

> Acho que na época, 2006, 2007 a Secretaria de Educação mesmo não estava, por exemplo, tão acostumada com o termo avaliação, com o uso de avaliações para orientar suas ações. Com o estabelecimento do SIMAVE, e aí avaliando as turmas eles perceberam que sim, que a informação derivada de uma avaliação externa pode trazer um conhecimento que não é explícito, que não é comum, que não é fácil de perceber e orientar o trabalho. Hoje em dia, a Secretaria de Educação lida muito melhor com a avaliação. Por exemplo, também teve resistência para acordo de resultados, que não é avaliação de política pública, mas está vinculado a monitoramento de resultados, então, tem que estabelecer metas. Hoje eu considero que a Secretaria de Educação está desenhando as próprias metas sem intervenção da SEPLAG ou de outro órgão porque eles entenderam a necessidade disso, quais são os obstáculos que estão postos para fazer metas desafiadoras e factíveis. Então eles internalizaram um entendimento que era esperado. Eu acho que a maior aprendizagem que teve, foi para a equipe que participou dessas avaliações dentro do Governo, foi de que o Governo não estava preparado para conduzir avaliações. Ele já era muito bom em monitoramento, mas ele não estava consciente dos custos ou dos processos dos quais ele teria que passar para poder aproveitar as avaliações de impacto para reorientar o seu caminho. (TÉCNICO EPR3, grifo da autora da dissertação)

> Quando a gente estava lá as pessoas começaram a entender. Na Superintendência de Informações Educacionais e nos gestores dos quatro programas. Eles começam a perceber o que é importante e o que não é importante. Porque a gente precisa da base de dados, faz as perguntas para o cara, mostra para o cara "Olha, não adianta você falar que tem tantos mil sendo atendidos porque outras coisas mudaram também no

> Brasil inteiro". Então você vai fazendo o cara entender o que é. O que o Governo talvez pudesse ser feito era ampliar isso ou mantido essa coisa e a coisa da base de dados, aí você vai gerando processo, vai entender os processos. Mesmo que o movimento que morreu, você começou a desenvolver uma capacidade do tema, da importância do tema chegando no nível gerencial. Você criou um conjunto de informações que estão em algum lugar. (COORDENAÇÃO CEDEPLAR, grifo da autora da dissertação)

> Eu acho que foi bom. E acho que até bom que teve, porque até se fomenta a fazer mais. Talvez se não tivesse sido feito na época, hoje a gente não estava nem pensando em fazer uma do Professor da Família. Alguém tem que iniciar as coisas, alguém tem que dar o ponta pé inicial. Mas é difícil convencer o pessoal a fazer. Agora, a gente está tentando agora, nesse governo, fazer algumas iniciativas de avaliação de impacto. Partindo a educação. A gente está com uma iniciativa, inclusive a gente a até já contatou o pessoal do CEDEPLAR também. Nós tivemos uma reunião com eles na semana passada, e a gente está com ideia de fazer uma avaliação de impacto para o Professor da Família. (EQUIPE SEE2, grifo da autora da dissertação)

Nesse sentido, é possível verificar, a partir dos relatos, a presença do uso processual (PATTON, 2008), ou seja, mudanças de comportamento que são reflexo da aprendizagem experimental adquirida ao longo da realização da avaliação e da convivência com avaliadores, apreendendo seus valores e forma de raciocínio. Outro efeito indireto positivo, também relacionado à internalização dos conhecimentos sobre avaliação por parte da SEE, diz respeito à melhoria da qualidade dos dados administrativos e registro dos projetos.

> Porque, na realidade, assim, os registros teriam que melhorar, os registros dos gerentes em relação aos critérios teriam que ter maior controle. Eu acho até que com isso deu uma melhorada. Uma vez que tenha sido um efeito indireto, não pensado, mas que teve. (EQUIPE SEE1)

Cabe refletir, no entanto, sobre o custo *versus* o benefício gerado por essas avaliações, como argumentado pelos entrevistados do EpR. Ou seja, a pergunta que fica, após a contratação de instituição externa de pesquisa, realização de treinamentos, contratação de funcionários e pesquisadores, criação de novas atividades tanto no EpR, como na SEE, entre outros, é: será que vale o tanto que se investiu? Na visão dos entrevistados do EpR, não. Para o técnico EpR3, as avaliações talvez tenham custado demais para gerar um resultado tão ínfimo, já que o investimento em mão de obra especializada feito foi desperdiçado devido a não continuidade do modelo. Nas palavras do entrevistado, *no governo de Minas Gerais o sistema de monitoramento e avaliação ficou cambota. A perna dele de monitoramento é mais forte do que a de avaliação, que é bem fraquinha, não tem bengalas para se apoiar*, como afirmado no trecho de entrevista abaixo.

> Essa frustração acho que é ruim. Tinha muita gente boa que trabalhava com isso e que ficou desanimado e desestimulado a trabalhar com isso dentro do governo, e aí é fogo, porque você fica com uma mão de obra muito qualificada fora do governo, no Cedeplar, na UFMG, na Fundação João Pinheiro, que foi até preparada para fazer essas avaliações, já que, na Fundação, você tinha um consultor externo sempre como referência, pessoas altamente gabaritadas. Depois, vai aproveitar a qualificação deles em que se não tem outras avaliações sendo conduzidas de forma organizada, centralizada e distribuída entre as secretarias ou, pelo menos, entre os projetos mais importantes? Aí a minha avaliação talvez tenha custado caro demais por ter feito tão pouco. Foi bastante frustrante, foi negativo, que não conseguiu, ficou fraco ou não foi forte o suficiente, pelo menos para poder internalizar esse sentimento de que é preciso avaliar as políticas públicas, de que é preciso pensar nas avaliações desde o início dos projetos. Porque na carteira da atual gestão também não foi pensado num plano de avaliação que fizesse desde o início o estabelecimento de quais eram os dados que precisariam ser coletados para que fosse possível num período de dois anos ou de quatro anos fazer uma avaliação para saber se aquele projeto mereceria mais recursos ou se ele já estaria pronto para ser encerrado por ter concluído com o seu objetivo ou por não está dando certo. Acho que no governo de Minas Gerais o sistema de monitoramento e avaliação ficou cambota. A perna dele de monitoramento é mais forte do que a de avaliação, que é bem fraquinha, não tem bengalas para se apoiar. (TÉCNICO EPR3, grifo da autora da dissertação)

Na visão do entrevistado da coordenação, a descontinuidade do modelo não permitiu que o trabalho gerasse *payback*, ou seja, dada a interrupção do desenvolvimento de mais avaliações não permitiu que o tempo decorrido entre o investimento inicial e o momento no qual o *lucro* – aqui entendido como o desenvolvimento dos projetos ou a aplicação das recomendações obtidas nas avaliações – acumulado se iguala ao valor desse investimento fosse alcançado. De acordo com o entrevistado, para que isso acontecesse, a avaliação teria que ser entendida como um processo, em que se desenha a política, implementa, monitora, avalia, faz as modificações necessárias e depois inicia de novo, como um ciclo que não deve ser interrompido. Como exemplificam os entrevistados.

> A pergunta é: vale o tanto que se pagou? Eu acho que não valeu, porque aconteceu o que sempre aconteceu, ou seja, foi um movimento que não teve continuidade, e nesse sentido não gerou *payback*. Por quê? Porque não chegou no ponto de dar *payback*. Então, você fez investimentos e ia ter um retorno em tanto tempo, você abortou a atividade de antes e você não pagou o investimento todo. Por quê? Porque você teve desrruptura institucional. E isso é o que aconteceu em vários momentos do governo. Do governo assim, da história de governos e tem vários exemplos nesse sentido. Seria [avaliação como processo], mas o processo tem que ter o ciclo dele, tem que desenhar, aí você implementa e monitora, avalia e muda de novo. É assim, o tempo inteiro fazendo pequenas correções. Você pode até fazer algumas, mas elas não podem mudar a linha da avaliação senão você perde a avaliação, e aí você não sabe se faz diferença dar um real ou não dar um real para o aluno para ele passar de ano ou não. De novo, tem coisa que você pode mexer e você deve mexer, tem cosias que não, mas isso tem que está no DNA do processo e da equipe. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

Esse sentimento de frustração demonstrado pelos entrevistados do EpR, dada a descontinuidade do modelo, está relacionado, de acordo com o entrevistado da coordenação,

com a imaturidade do governo para realizar avaliações de políticas públicas. Para o entrevistado, o exercício de retroalimentar programas governamentais a partir de informações técnicas contidas nas avaliações demanda a criação de um ambiente propício a essa atividade, o que o governo ainda não estaria preparado para desenvolver, como mostrado no trecho de entrevista abaixo.

> Mas, enfim, qual que é a variável que coloca isso? A maturidade do governo. O governo ainda não tem maturidade e maturidade demanda tempo. Ou seja, você teria que ter uma instituição com um pouco mais de tempo, alguns anos a mais para entender e ter um ambiente propício para você fazer uma opção de um retorno mais em longo prazo do que um retorno mais imediato desenvolvendo mais capacidade. Mas, daqui há alguns anos, a gente pode ver isso acontecendo. Tem a ver quando você pega a experiência de Inglaterra e outros países com algumas décadas a mais que a nossa instituição democrática você vê que isso acontece. (COORDENAÇÃO EPR1, grifo da autora da dissertação)

Por fim, como conclusão final da avaliação do projeto PAV, observa-se, infelizmente, uma não utilização completa dos resultados, já que relatos apontaram para a implantação de mudanças no projeto antes mesmo que a avaliação fosse concluída, o que assinala para um processo de avaliação ignorado pelos próprios usuários. Dentre os projetos aqui estudados, verifica-se maior resistência por parte dos gestores devida, em parte, ao entendimento que o cliente da avaliação era o EpR, ao contrário da SEE, e até pela falta de familiaridade, no início da avaliação, com métodos de avaliação. Por outro lado, nota-se o desenvolvimento de capacidade de avaliação, tanto na instituição de pesquisa contratada, que adquiriu mais experiência no uso de dados secundários para avaliações de impacto, quanto na SEE, que hoje detém conhecimentos em avaliação e até iniciativa própria de demandar suas avaliações.

11 CONSIDERAÇÕES FINAIS

As avaliações de políticas públicas carregam em si um receituário de boas intenções. As promessas vão, desde a contribuição para aumentar a racionalidade na tomada de decisões, identificando problemas, selecionando alternativas de solução e suas consequências, até o processo de aprendizagem organizacional. Constituem também peças fundamentais das boas práticas da administração pública, na medida em que instituem valores de eficiência e democracia. De eficiência, ao demonstrar uma preocupação clara com a questão dos resultados das políticas públicas, definem um importante papel para um rico e dinâmico conjunto de abordagens avaliatórias focadas na efetividade, eficácia, eficiência de programas governamentais. Com relação à democracia, as avaliações permitem que os governos tenham melhores evidências para a implementação de programas governamentais e para aqueles que já estão em andamento, buscando implementar intervenções com maior benefício social para a população. No entanto, apesar de ser clara a percepção de que a prática da avaliação de programas governamentais pode levar a uma transformação nas estruturas de conhecimento de organizações públicas, ou seja, na forma como ela detecta e corrige seus erros, a grande questão que surge nesse espaço é: as avaliações são realmente úteis ou constituem apensas mais um relatório amontoado nas gavetas dos gestores público?

O objetivo geral deste trabalho consistiu em *analisar a dinâmica da utilização dos resultados das avaliações de Projetos Estruturadores realizadas pelo Governo do Estado de Minas Gerais a partir do ano de 2007*. Em termos específicos, buscou-se: 1) identificar e analisar quais os fatores e mecanismos que influenciam a utilização (ou não) dos resultados das avaliações; 2) analisar o contexto e a conjuntura dos processos de implementação de avaliação de programas governamentais no governo do estado de Minas Gerais e quais são os principais resultados já identificados; 3) analisar como os instrumentos de avaliação têm contribuído para o aprimoramento e introdução de mudanças nas políticas avaliadas no contexto do governo do estado de Minas Gerais; 4) compreender como as abordagens adotadas nas avaliações de políticas públicas realizadas pelo governo do estado de Minas Gerais influenciaram a utilização de seus resultados; 5) propor alternativas de procedimentos para a utilização mais efetiva das avaliações dos programas governamentais.

Como se verificou no presente estudo, a partir da discussão teórica e da análise empírica do caso de Minas Gerais, embora exista um crescente interesse em avaliações de programas governamentais no Brasil e nos governos estaduais, pouca atenção tem sido dada à utilização das informações e resultados de programas governamentais. Além da escassez de estudos empíricos nacionais que permitam verificar a utilização de tais avaliações como instrumento de correção de rumos e aprimoramento dos serviços públicos, constatou-se que há uma série de desafios à realização concreta dessa concepção normativa de avaliação de políticas públicas, o que se buscará consolidar nesta conclusão.

Um primeiro ponto a ressaltar se refere às várias facetas de uma avaliação. Embora representem estudos técnicos que têm como finalidade identificar obstáculos, propor medidas de correção e alteração de políticas e programas a partir de evidências, as avaliações não constituem atividades destituídas de valores, como se fossem instrumentos técnicos imparciais ou impessoais. Nesse sentido, não se pode entender a avaliação como uma ferramenta meramente técnica, ao contrário, como afirmado por Draibe (2001), é uma atividade que, por sua própria natureza, condições e métodos, constitui uma ação conflitiva que consolida valores, afirma interesses, provoca mudanças e transforma pensamentos. Independentemente de qual seja a proposta de uma dada avaliação, seu desenvolvimento sempre irá apoiar-se em uma estrutura organizacional já definida e em um sistema gerencial e decisório que podem influenciar as suas condução e implementação.

Logo, a premissa de que os resultados das avaliações serão, direta e automaticamente, utilizados para promover as melhorias necessárias foi aqui desmistificada. Percebeu-se que a existência de avaliações por si só não garante o melhoramento da gestão pública e que o uso apropriado dos resultados das avaliações não constitui tarefa trivial, já que a utilização das avaliações representa um fenômeno caracterizado por fatores como relevância da avaliação, credibilidade do avaliador, envolvimento com o usuário, comunicação, atendimento às necessidades de informação, percepção da avaliação como instrumento de gerenciamento e, principalmente, características do contexto de uma instituição. A utilização pode ser vista como um fenômeno complexo aliado ao contexto, em que pessoas, ambiente e circunstâncias estão fortemente conectados, influenciando o uso.

Quanto às avaliações dos projetos estruturadores aqui estudadas, observou-se, a partir das evidências geradas com a realização das entrevistas, que as decisões relativas à

implementação de melhorias no âmbito dos programas, tanto no âmbito gerencial, quanto na esfera política, pouco foram influenciadas pelos resultados das avaliações. Dos quatro casos estudados, os únicos que apresentaram uma provável utilização instrumental, ou seja, uso relativo ao apoio às decisões e à busca de resolução de problemas a partir da avaliação, foram as avaliações dos projetos Gestão Integrada de Informações de Defesa Social, cujo objeto avaliado foi a redução da criminalidade nas áreas em que o projeto IGESP havia sido implantado, e o projeto Viva Viva, no qual se avaliou o impacto dos Centros Viva Vida na redução da mortalidade materna. No primeiro caso, verificou-se um uso legitimativo relacionado aos resultados positivos da avaliação de ampliar a metodologia IGESP para as demais regiões de gestão integrada do estado de Minas Gerais, mas não se sabe ao certo se essa decisão já seria implementada independentemente da realização do estudo avaliatório. No caso do Viva Vida, alguns avaliadores inferem que os problemas vivenciados ao longo da avaliação, como a dificuldade de obter um registro confiável das gestantes de risco, culminaram na criação de um novo projeto, o Mães de Minas. Observa-se, contudo, que, em ambos os projetos, as percepções e a aplicação da avaliação variam muito e não foi possível encontrar um consenso entre os grupos de *stakeholders* avaliados.

Em relação aos projetos PAV e Minas Sem Fome, não foi possível verificar uma modificação nos objetos avaliados derivada das avaliações. No caso do Minas Sem Fome, concluiu-se pelo uso conceitual, quando a avaliação tem capacidade de influenciar a opinião dos técnicos a respeito de um determinado programa, sem que exista qualquer ação em decorrência dos resultados de uma avaliação. Já, no caso do PAV, no qual se observou um não uso completo, relatos apontaram para a implantação de mudanças no projeto antes mesmo que a avaliação fosse concluída, o que sugere um processo de avaliação ignorado pelos próprios usuários.

Em parte, ainda que muitas percepções dos entrevistados pareçam difusas, acredita-se que as entrevistas aqui analisadas retrataram aspectos significativos do modelo de gestão pública mineiro, evidenciando anseios e expectativas de diferentes grupos envolvidos nas avaliações. Ao analisar os dados obtidos nas diferentes organizações pesquisadas, foi possível identificar uma série de circunstâncias, algumas específicas de cada contexto avaliado, outras que são comuns a todos os estudos avaliatórios, que, de alguma forma, influenciaram positiva ou negativamente a utilização de seus resultados.

Em correspondência com a teoria pesquisada, uma das possíveis características das avaliações do modelo mineiro, que podem ter comprometido o uso, está relacionada com o desenho *top-down* das avaliações que também repercute no grau de envolvimento does gestores e equipes dos projetos avaliados. A constituição do EpR e da FJP como *sponsors* das avaliações, responsabilizando-se pelos processos de contratação e definição dos projetos avaliados, causou uma anomalia quanto `a definição de quais seriam os clientes principais dos resultados dos estudos. No entendimento dos avaliadores e dos próprios gestores, ao contrário de eles mesmos se verem como interessados diretos nas avaliações, o EpR, por ser o demandante do processo, também seria o cliente da avaliação ainda que esse não detivesse controle direto sobre os projetos.

Tal distorção quanto à consideração de quais seriam os usuários primários da avaliação pode ter criado, além de resistências internas dado o caráter imposto de realização, avaliações que não estão alinhadas com as necessidades de informação dos *stakeholders* diretamente envolvidos na gestão dos projetos, que pode ser ilustrada pelas participações reduzidas dos gestores e equipe dos projetos nos processos de implementação e execução das avaliações. Em geral, verificou-se um papel mais processual nas unidades setoriais, limitando-se ao fornecimento de informações sobre o projeto e dados, quando solicitados pelo próprio EpR, que atuava como intermediário entre avaliadores e equipe do projeto avaliado. Devido a isso, alguns usuários reais demonstraram indiferença quanto aos possíveis resultados das avaliações.

Outro motivo relacionado à falta de dedicação dos gestores dos projetos com as avaliações diz respeito à priorização da rotina e das atividades diárias necessárias à execução do projeto, em detrimento da gestão de longo prazo, focada nos resultados e nos efeitos das ações implementadas. A falta de tempo para se dedicar `a implementação das ações foi mencionada nos casos das avaliações dos projetos Viva Vida, Gestão Integrada e PAV o que aborda o que Mokate (2002) chama de foco em atividades, um estilo de trabalho orientado para o desempenho de rotinas de trabalho predefinidas, em que as atividades tornam-se fins em si mesmas. Dessa forma, a equipe de trabalho é orientada por atividades de conformidade e processos, em que a máxima desse estilo de trabalho é a entrega de um bem ou serviço, o que requer apenas informações de monitoramento de processos, não existindo, assim, interesse e/ou incentivos para realizar ou para solicitar uma avaliação dos resultados ou dos impactos de um determinado projeto.

Os avaliadores dos projetos, por sua vez, também focados em atividades de execução da avaliação, principalmente com as questões metodológicas, não tecem esforços para inserir as equipes gestoras dos projetos no desenvolvimento dos estudos avaliatórios ou, até mesmo, após a sua conclusão. Observou-se que o entendimento dos avaliados quanto ao período de participação nas avaliações vai até o término e entrega dos resultados aos demandantes do trabalho, sejam eles usuários ou não. Consideram, assim, que a função dos avaliadores resume-se em entregar resultados, quando, na verdade, deveria conjugar ações de inclusão, geração de capacidade e subsídio na utilização das descobertas dos dados por eles encontradas. Tal percepção limitadora do papel do avaliador pode ser comprovada pelo desconhecimento geral, demonstrado por todos os dez avaliadores entrevistados, sobre o uso das avaliações por eles desenvolvidas. Em nenhuma das avaliações aqui pesquisadas identificou-se a criação de incentivos para a sua utilização e entendimento das mesmas, pelo contrário, havia uma clara orientação das equipes externas pela conclusão.

Em relação às questões metodológicas, tanto ligadas aos processos percorridos pelos estudos avaliatórios examinados, quanto à natureza das avaliações (avaliações de impacto), verifica-se uma série de fatores que podem ter interferência no seu uso. Em princípio, tratando-se das dificuldades encontradas no desenvolvimento dos processos da avaliação, observou-se, de forma unânime, a precariedade do desenho das políticas públicas que repercute na falta de dados e informações consolidadas sobre os projetos avaliados. A ausência de dados consolidados e confiáveis remete ao fato de que as políticas são implementadas sem que incorporem as necessidades de informação do projeto, seja para uma futura avaliação seja para controle interno, no momento em que são formuladas. Percebeu-se, assim, uma necessidade clara de fortalecer e aprimorar os sistemas de informações gerenciais, com dados de desempenho dos projetos, a fim de conduzir com proveito o monitoramento dos projetos, criar indicadores de desempenho de médio prazo e/ou desenvolver possíveis estudos de avaliação, que, por sua vez, já devem ser planejados concomitantemente `a fase de desenho dos projetos.

Já, quanto ao tipo de avaliação selecionado, observam-se peculiaridades das avaliações de impacto que são pouco direcionadas ao uso. Um primeiro ponto se refere à complexidade inerente às avaliações de impacto dada a necessidade de utilizar métodos que permitam a aferição das relações de causa e efeito das políticas públicas. Para estimar o efeito causal de

um programa sobre seu público-alvo, o método escolhido, sendo o aleatório o método mais indicado, deve simular o que aconteceria com os beneficiários caso o programa não tivesse ocorrido, com o objetivo de descobrir quais mudanças nos beneficiários são diretamente atribuíveis ao programa em questão. A sofisticação dos métodos, inerente a esse tipo de avaliações, pode, contudo, distanciar o gestor dos projetos da prática de avaliação. Um segundo ponto se refere à realização mais comum de estudos não aleatorizados, o que implica conclusões insuficientes e, até, controversas, que colocam em cheque a real utilidade das avaliações. Por fim, a constatação de que as avaliações de impacto contemplam ciclos de desenvolvimento longos o que nem sempre é compatível com o tempo político do governo. Prova disso foi a elevada rotatividade citada nas entrevistas das pessoas envolvidas com as avaliações de impacto aqui pesquisadas, tanto por parte dos gestores dos projetos, quanto dos avaliadores, já que, como foi mostrado, um avaliador que iniciava a avaliação não era o mesmo que concluía ou o gestor que recebeu a avaliação não era o mesmo que participou da apresentação dos resultados. Além disso, uma das dificuldades dessa pesquisa foi justamente encontrar potenciais entrevistados para desta pesquisa, já que muitos não se encontravam na mesma função de quando as avaliações foram iniciadas.

Torna-se importante também elencar os motivos que influenciaram o uso das avaliações desenvolvidas, o balanço entre a pretensão da iniciativa e a maturidade do governo para absorver as avaliações. É necessário refletir sobre os paradigmas existentes no setor público, como o ambiente político nos quais os programas se desenvolvem, a cultura administrativo-financeira que impõe a determinação sobre a alocação do gasto público, além da tradição da decisão baseada no conhecimento prático das pessoas envolvidas nos programas que colocam valores e interesses frente a evidências, que concorrem fortemente com a utilização das avaliações de forma técnica e racional. Interpor um sistema de avaliações conciso, que tem como objetivo superar todos esses fatores mencionados, como foi a proposta do governo de Minas Gerais, por si só já se caracteriza como um projeto ousado.

Nesse sentido, ressaltam-se nesta oportunidade, quatro pontos comuns observados nas avaliações de políticas aqui pesquisadas que podem, em maior ou menor grau, ter influenciado o uso precário das recomendações e conclusões encontradas.

1) Ausência de orientação quanto às necessidades de informação por parte dos usuários reais das avaliações, uma vez que as avaliações foram conduzidas de

acordo com a lógica dos avaliadores e não são direcionadas às necessidades de informação e prazo de seus *stakeholders* fundamentais.

2) Falta de tempo dos gestores dos projetos avaliados para lidar com as demandas de avaliação colocadas, dado que os usuários são consumidos pelas atividades de entregas de rotina do projeto em detrimento das preocupações com os resultados desempenhados.

3) Dificuldade de atribuir causalidade aos programas avaliados, seja por falta de dados completos, seja pela ausência de um planejamento da avaliação concomitante à formulação do projeto.

4) Insuficiência de experiências em conduzir, revisar e utilizar as avaliações, devido à cultura da administração pública que não prioriza o uso de evidências, como também à carência de familiaridade com os métodos de avaliação que limitam o potencial de uso das avaliações.

Contudo, ainda com todas as limitações e constrangimentos inerentes à realização de avaliações das políticas públicas, verificou-se um consenso entre os atores entrevistados sobre a importância dessa iniciativa para uma postura mais racional do setor público, seja mineiro, seja nacional. Nenhum entrevistado, por mais questionador quanto ao processo de avaliação, duvidou do valor inegável presente na realização de avaliações para a efetividade da administração pública em suas várias instâncias governamentais e da necessidade crescente de investimentos na área. Ganha destaque, portanto, o reconhecimento da possibilidade dos gestores públicos de dispor de processos de avaliação que lhes permitam controlar e articular ações de aprimoramento dos projetos a partir de evidências.

O que não pode acontecer, no entanto, é priorizar a difusão de uma *moda* de avaliações, sem que seu problema mais fundamental, a utilização dos resultados, esteja institucionalizado como uma prática intrínseca à realização. Assim sendo, no esforço de unir a teoria de avaliações estudada e as percepções dos estudos de caso selecionados, como também de atender ao objetivo específico desta dissertação de *propor alternativas de procedimentos para a utilização mais efetiva das avaliações dos programas governamentais*, indicam-se, aqui, algumas ações prioritárias que poderiam promover o uso mais apropriado das avaliações. Como recomendações destacam-se as seguintes:

- assegurar uma participação ativa dos reais usuários das avaliações na fase de discussão sobre a necessidade de se fazer uma avaliação e do seu desenho, garantindo que seja construída com o apoio e entendimento geral dos clientes, em termos de necessidades de informação, funcionamento do programa e o contexto em que a avaliação é implementada;
- garantir a familiaridade dos usuários com os instrumentos de avaliação e os pontos positivos e negativos presentes em cada uma das abordagens avaliatórias, convidando-os, quando pertinente, a participar de programas de treinamento sobre o assunto;
- selecionar métodos adequados às perspectivas reais de informação dos programas avaliados e incluir estratégias de avaliações que unam as abordagens qualitativas às quantitativas, com foco não só nos efeitos das políticas, mas também no desenho e implementação dos projetos;
- manter os usuários envolvidos e informados durante todas as etapas da avaliação e não só na etapa de conclusão do relatório, com insumos sobre o progresso, as dificuldades e primeiras hipóteses do trabalho, de forma a garantir que não haja grandes surpresas quanto ao resultado final da avaliação;
- entender como os usuários gostariam de receber as informações e garantir que as apresentações estejam adequadas aos diferentes tipos de audiência;
- desenvolver e monitorar um plano de ação das recomendações plausíveis das avaliações com o auxílio de todos os *stakeholders* envolvidos e de acordo com as possibilidades e contexto do momento em que as avaliações são concluídas de forma a certificar-se da continuidade e a efetividade do estudo.

As recomendações aqui sugeridas não têm qualquer pretensão de criar uma lista de passos para o projeto de avaliações úteis, mas, sim, evidenciar desafios e etapas importantes que ainda são esquecidas nas avaliações aqui pesquisadas e que se acredita que também sejam negligenciadas em muitas das avaliações realizadas no Brasil.

Por fim, como consideração final deste trabalho, espera-se que, com os resultados e as considerações apresentados, tenha-se contribuído para o entendimento dos aspectos relacionados ao uso das avaliações de programas governamentais ainda pouco explorados no Brasil. Cabe almejar também que esta dissertação registre o exercício de reflexão feito sobre a

dinâmica da efetividade das avaliações realizadas em governos para que, de alguma forma, elucide o desperdício de recursos públicos investidos em avaliações que *vão para a gaveta*. Isto posto, para possíveis estudos desenvolvidos nessa área, acredita-se que os argumentos aqui construídos e as questões aqui evocadas possam elucidar novas perguntas relativas ao tema das avaliações de políticas públicas, buscar recortes mais específicos ou novas possibilidades de encadeamento dos conceitos e ideias. O que não se esgota é o tema pesquisado.

REFERÊNCIAS

AGUILAR, M. J.; ANDER-EGG, E. **Avaliação de serviços e programas sociais**. Rio de Janeiro: Vozes, 1995.

ALA-HARJA, M.; HELGASON, S. Em direção às melhores práticas de avaliação. **Revista do Serviço Público**, Brasília, v. 51, n. 4, p. 5-59, out.dez. 2000.

ALKIN, M. C. **Evaluation roots**: tracing theorists views and influences. Thousand Oaks, California: Sage, 2004.

ALKIN, M. C.; COYLE, K. Thoughts on evaluation utilization, misutilization, and non-utilization. **Studies in Educational Evaluation**, v. 14, p. 331-40, 1988.

ALMEIDA, P. R. de. A experiência brasileira em planejamento econômico: uma síntese histórica, 2006. Disponível em: <http://www.pralmeida.org/05DocsPRA/1277HistorPlanejBrasil.pdf>. Acesso em: 05 dez. 2011.

ANASTASIA, A. A. J. Apresentação. In: MARINI, C. *et. al.* (Org) **O choque de gestão em Minas Gerais**: políticas da gestão pública para o desenvolvimento. Belo Horizonte: Editora UFMG, 2006, p. 13-17.

ARRETCHE, M. T. S. Tendências no estudo sobre avaliação. In: RICO, E. M. (Org). **Avaliação de políticas sociais**: uma questão em debate. São Paulo: Cortez, 1998, p. 29-39.

BAMBERGER, M.; RUGH, J.; MABRY, L. **Real world evaluation**: working under budget, time, data, and political constraints. Thousand Oaks: Sage, 2011.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 1988.

BECHELAINE, C. H. de O. **A gestão pública por resultados no estado de Minas Gerais: um estudo de caso sobre o Programa Estado para Resultados**. Monografia (Graduação em Administração Pública) – Fundação João Pinheiro, Escola de Governo de Minas Gerais. Belo Horizonte, 2009.

BERNARDI, M. M. E.; GUIMARÃES, T. B. Empreendedores públicos: uma experiência de gestão estratégica de pessoas na administração pública do Governo do Estado de Minas Gerais. In: CONGRESO INTERNACIONAL DEL CLAD SOBRE LA REFORMA DEL ESTADO Y DE LA ADMINISTRACIÓN PÚBLICA, 13., 2008, Buenos Aires. **Anais...** Buenos Aires: CLAD, 2008.

BREMNER, R. **From the depths**: the discovery of poverty in America. New York: New York University, 1956.

BUSJEET, G. D. B. Transformando o setor público: a introdução do M&A em Minas Gerais. In: GUIMARÃES; T. B.; PORDEUS; I. A; CAMPOS, E. S. A. (Orgs). **Estado para Resultados**: avanços no monitoramento e avaliação da gestão pública em Minas Gerais. Belo Horizonte: UFMG, 2010, p. 175-206.

CALMON, K. M. N. A avaliação de programas e a dinâmica da aprendizagem organizacional. **Revista de Planejamento e Políticas Públicas**, Brasília, n. 19, p. 1-70, jun. 1999.

CALMON, K.; GUSSO, D. A. A experiência de avaliação do Plano Plurianual (PPA) do Governo Federal no Brasil. **Planejamento e Políticas Públicas**, Brasília, n. 25, jun 2002.

CARNEIRO, R.; BARBOSA, L. O. S. Acordo de Resultados: Análise da Experiência do Governo Mineiro. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 31., 2007, Rio de Janeiro. **Anais...** Rio de Janeiro: ANPAD, 2007.

CASTRO, M. H. G. de. **Avaliação de políticas públicas e programas sociais**. Campinas: UNICAMP/NEPP, 1989. (Caderno de Pesquisa, n. 12).

COHEN, E.; FRANCO, R. **Avaliação de projetos sociais**. Petrópolis: Vozes, 1993.

CORRÊA, I. M. Planejamento estratégico e gestão pública por resultados no processo de reforma administrativa do Estado de Minas Gerais. **Revista de Administração Pública**, Rio de Janeiro, n. 41, v. 3, p. 487-504, mai,jun. 2007.

COSTA, F. L. da; CASTANHAR, J. C. Avaliação de programas públicos: desafios conceituais e metodológicos. **Revista de Administração Pública**, Rio de Janeiro, n. 37, v. 5, p. 969-2, mai. 2003.

COUSINS, J. B.; LEITHWOOD, K. A. Current Emprirical research on evaluation utilization. **Review of Educational research**, v. 56, n. 3, p. 331-64, fall 1986.

COX, G. B. Managerial style: implications for the utilization of program evaluation information. **Evaluation,** v. 1, n. 4, p. 499-508, dec. 1977.

DAHLER-LARSEN, P. Evaluation and Public Management. In: FERLIE, E.; LYNN Jr, L.; POLLITT; C (Orgs). **The Oxford handbook of public management**. Oxford: Oxford University, 2005, p. 615-39.

DRAIBE, S. M. Avaliação de implementação: esboço de uma metodologia de trabalho em políticas públicas. In: BARREIRA, M. C. R. N.; CARVALHO, M. C. B. de (Orgs). **Tendências e perspectivas na avaliação de políticas e programas sociais**. São Paulo: IEE-PUC-SP, 2001, p. 13-42.

DUARTE, K.; LEMOS, C.; MARINI, C.; MARTINS, H. F. Acordo de resultados. In: MARINI, C. *et. al.* (Orgs) **O choque de gestão em Minas Gerais**: políticas da gestão pública para o desenvolvimento. Belo Horizonte: Editora UFMG, 2006, p. 95- 127.

FARIA, C. A. P. A Política da avaliação de políticas públicas. **Revista Brasileira de Ciências Sociais**, São Paulo, v. 20, n. 59, p. 97-169, out. 2005.

FERREIRA, A. H. B.; LOPES, H. M.; CAMPOS, E. S. A. Avaliação de políticas públicas: o caminho percorrido por Minas Gerais. In: GUIMARÃES; T. B.; PORDEUS; I. A; CAMPOS,

E. S. A. (Orgs). **Estado para resultados**: avanços no monitoramento e avaliação da gestão pública em Minas Gerais. Belo Horizonte: UFMG, 2010, p. 96-112.

FIGUEIREDO, M. F.; FIGUEIREDO, A. M. C. Avaliação política e avaliação de políticas: um quadro de referência teórica. **Análise e Conjuntura,** Belo Horizonte, v. 1, n. 3, p. 107-27, set.dez. 1986.

FUNDAÇÃO JOÃO PINHEIRO. Centro de Estudos de Políticas Públicas. Avaliação Programa de redução da mortalidade infantil e materna Viva Vida. / Centro de Estudos de Políticas Públicas da Fundação João Pinheiro. – Belo Horizonte, 2010a.

FUNDAÇÃO JOÃO PINHEIRO. Centro de Estudos de Políticas Públicas. Avaliação de Impacto da Ação de Lavoura do Programa Minas Sem Fome. / Fundação João Pinheiro. Centro de Estudos de Políticas Públicas. Belo Horizonte, 2010b.

FUNDAÇÃO JOÃO PINHEIRO. Centro de Estudos de Políticas Públicas. Integração das Polícias Civil e Militar: uma avaliação da experiência de Minas Gerais no âmbito do programa "Integração da Gestão em Segurança Pública" (IGESP). / Fundação João Pinheiro. Centro de Estudos de Políticas Públicas. Belo Horizonte, 2010c.

FUNDAÇÃO JOÃO PINHEIRO. Centro de Estudos de Políticas Públicas. Viabilidade econômica financeira do programa Minas Sem Fome, Ações de Tanque de Resfriamento de Leite e Apicultura. / Fundação João Pinheiro. Centro de Estudos de Políticas Públicas. Belo Horizonte, 2010d.

GARCES, A.; SILVEIRA, J. P. Gestão pública orientada para resultados no Brasil. **Revista do Serviço Público**, Brasília, v. 53, n. 4, p. 53-76, out.dez. 2002.

GARCIA, R. C. **Subsídios para organizar avaliações da ação governamental**. Brasília, DF: IPEA, 2001, 68 p. (Texto para discussão, n.776).

GERTLER, P. J. et al. **Impact evaluation in practice**. Washington, DC: The World Bank, 2011.

GUBA, E. G. Problems in utilizing the results of evaluation. **Journal of Research and Development in Education**, Athens, v. 8, n. 3, p. 42-54, spring 1975.

GUBA, E. G.; LINCOLN, Y. S. **Fourth generation evaluation**. Newbury Park: Sage, 1989.

GUIMARÃES, T. B.; ALMEIDA, B. de. O duplo planejamento. In: MARINI, C. *et. al.* (Orgs) **O choque de gestão em Minas Gerais**: políticas da gestão pública para o desenvolvimento. Belo Horizonte: Editora UFMG, 2006, p. 43-54.

GUIMARÃES, T. B.; CAMPOS, E. S.. A. Gestão da estratégia no governo do estado de Minas Gerais: em busca da harmonia entre a ação e os resultados. CONGRESSO INTERNACIONAL DEL CLAD SOBRE LA REFORMA DEL ESTADO Y DE LA ADMINISTRACIÓN PÚBLICA, 13., 2008, Buenos Aires. **Anais...** Buenos Aires: CLAD, 2008.

GUIMARÃES, T. B., ; CAMPOS, E. S. Sistema de monitoramento e avaliação no governo do estado de Minas Gerais: aspectos de gerenciamento. 2009. Disponível em: < http://siteresources.worldbank.org/INTLACREGTOPPOVANA/Resources/COORDENAÇÃO EPR2_Barreto_Minas_Gerais.pdf> Acesso em: 14 out. 2010.

HAYES, S. P. Jr. **Evaluating development projects**. Paris: Unesco, 1959.

HENRY, G. T. Why not use? In: CARACELLI, V. J.; PRESKILL, H. (Eds) **The expanding scope of evaluation use**. San Francisco: Jossey-Bass, 2000, p. 85-98.

JOHNSON, R. B. Toward a theoretical model of evaluation utilization. **Evaluation and Planning**, v. 21, n. 1, p. 93-110, fev. 1998.

JOHNSON, K. *et al.* Research on evaluation use: a review of the empirical literature from 1986 to 2005. **American Journal of Evaluation**, Champaign, v. 30, n. 3, p. 377-410, aug. 2009.

JOINT COMMITTEE FOR EDUCATIONAL EVALUATION. Disponível em: <http://www.jcsee.org/about>. Acesso em: 05 dez. 2011.

KHANDKER, S.R., KOOLWAL, G.B., SAMAD, H.A. **Handbook on impact evaluation: quantitative methods and practices**. Washington, D.C: The World Bank, 2010.

LEVITON, L. C.; BORUCH, R. F. Contributions of evaluation to education programs and policy. **Evaluation Review**, Birmingham, v. 7, n. 5, p. 563-98, oct. 1983.

LEVITON, L. C.; HUGHES, E. F. X. Research on the utilization of evaluations: A review and synthesis. **Evaluation Review**, Birmingham, v. 5, n. 4, p. 525-48, aug. 1981.

LOBO, T. A avaliação de processos e impactos em programas sociais: algumas questões para reflexão. In: RICO, E. M. (Orgs). **Avaliação de políticas sociais**: uma questão em debate. São Paulo: Cortez, 2001.

MINAS GERAIS. Secretaria de Estado de Planejamento e Gestão. **Plano Mineiro de Desenvolvimento Integrado 2003-2020**. Belo Horizonte, 2003.

MINAS GERAIS. Secretaria de Estado de Planejamento e Gestão. **Plano Mineiro de Desenvolvimento Integrado 2007-2023**. Belo Horizonte, 2007.

MINAS GERAIS. Secretaria de Estado de Planejamento e Gestão. Subsecretaria de Planejamento e Orçamento. **Revisão do Plano Plurianual de Ação Governamental 2008-2011 – Exercício 2009**. Belo Horizonte, 2009.

MINAS GERAIS. Lei n. 17.007, de 28 de setembro de 2007. Atualiza o Plano Mineiro de Desenvolvimento Integrado (PMDI) e dá outras providências. Minas Gerais, Belo Horizonte, 2007. Disponível em: <http:// www.almg.gov.br>. Acesso em: 10 out. 2012.

MINAS GERAIS. Secretaria de Estado de Saúde. Mães de Minas, ávida merece esse cuidado. Belo Horizonte, 2012. Disponível em: < http://www.maesdeminas.com.br/>. Acesso em: 12 jan. 2013.

MINTZBERG, H. **The Nature of Managerial Work**. New York: Harper & Row, 1973.

MOKATE, K. M. Convirtiendo el "monstruo" en aliado: la evaluación como herramienta de la gerencia social. **Revista do Serviço Público**, Brasília, v. 8, n.1, p. 91-136, 2002.

NADER, M. R. A avaliação como ferramenta para uma gestão pública orientada para resultados: o caso do Governo Federal Brasileiro. In: Documento elaborado en el año 2005, por encargo del CLAD, 2005. Disponível em: <http://unpan1.un.org/intradoc/groups/public/documents/CLAD/clad0053502.pdf>. Acesso em: 30 ago. 2011.

OLIVEIRA, K. P. de. **Avaliação de implementação de políticas de recursos humanos no setor público**: o caso do governo de Minas Gerais. 2009. 229f. Dissertação (Mestrado em Administração)– Faculdade de Ciência Econômicas, Universidade Federal de Minas Gerais, Belo Horizonte.

OSPINA, S. Una aproximacion sistémica a la evaluacion de la gestion publica. In: **La evaluación como herramienta para una gestión pública orientada a resultados**. Caracas: Centro Latinoamericano de Administración para el Desarollo (CLAD), 2002. Disponível em: <http://www.clad.org/siare_isis/innotend/evaluacion/aproximacion.pdf>. Acesso em: 09 fev. 2012.

PATTON, M. Q. *et al.* In search of impact: an analysis of the utilization of federal health evaluation research. In: WEISS, C. H. (ed.) **Using social research in public policy making**. Lexington, MA: Lexington Books, 1977, p. 141-63.

PATTON, M. Q. **Utilization-focused evaluation**. Thousand Oaks: Sage, 2008.

PEREZ, J. R. R. A avaliação do processo de implementação: algumas questões metodológicas. In: RICO, E. M. (Org). **Avaliação de políticas sociais:** uma questão em debate. São Paulo: Cortez, 2001, p. 65-73.

PISCO, L. A. A avaliação como instrumento de mudança. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 11, n. 3, p. 564-76, set. 2006.

RIBEIRO, S. M. R. Avaliação no Brasil: questões de teoria e de prática. XIV CONGRESO INTERNACIONAL DEL CLAD SOBRE LA REFORMA DEL ESTADO Y DE LA ADMINISTRACIÓN PÚBLICA. Salvador, Bahia, 2009. **Anais...** Salvador: CLAD, 2009.

ROSEMBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, Oxford, v. 70, n. 1, p. 41-55, 1983.

ROSSI, P.; FREEMAN, H. **Evaluation:** a systematic approach. Newbury Park: Sage, 1993.

SAPORI, L. F.; ANDRADE, S. C. P. de. Integração policial em Minas Gerais: desafios da governança da política de segurança pública. **Civitas**, Porto Alegre, v. 8, n. 3, p. 428-53, set.dez. 2008.

SERPA, S. M. H. C. **Para que avaliar?** identificando a tipologia, os propósitos e a utilização das avaliações de programas governamentais no Brasil. 2010. 180f. Dissertação (Mestrado em Administração) – Universidade de Brasília, Brasília.

SILVA, P. L. B.; MELO, M. A. B. de. O processo de implementação de políticas públicas no Brasil: características e determinantes da avaliação de programas e projetos. Campinas: NEPP: Unicamp. (Caderno, 48), 2000.

SCRIVEN, M. S. An introduction to meta-evaluation. **Educational Products Report**, v. 2, n. 5 , p. 36-8, fev. 1969.

SHULHA, L. M.; COUSINS, J. Evaluation use: theory, research, and practice since 1986. **Evaluation Practice**, v. 18, n. 3, p. 195-209, 1997.

SOARES, R. R.; SOUZA, I. V.M. Organization and information in the fight against crime: an evaluation of the integration of police forces in the state of Minas Gerais, Brazil. In: 32° Meeting of the Brazilian Econometric Society, 2010, Salvador. **Anais...** Salvador: SBE, 2010.

STUFFLEBEAM, D. L. The metaevaluation imperative. **American Journal of Evaluation**, Champaign, v. 22, n. 2, p. 183-209, 2001.

SUCHMAN, E. A. **Evaluative research; principles and practice in public service & social action programs**. New York: Russell Sage Foundation, 1967.

VERGARA, S. C. **Métodos de pesquisa em administração**. São Paulo: Atlas, 2006.

VILHENA, R. Rumo à segunda geração do choque de gestão. In: MARINI, C. *et. al.* (Org.) **O choque de gestão em Minas Gerais:** políticas da gestão pública para o desenvolvimento. Belo Horizonte: Editora UFMG, 2006, p. 351-56.

WORTHEN B. R.; SANDERS J.R.; FITZPATRICK J. L. **Avaliação de programas:** concepções e práticas. São Paulo: Gente, 2004.

WEISS, C. H. **Utilization of evaluation**: toward comparative study. Bostton, MA: Allyn and Bacon, 1972.

WEISS, C. H. Improving the linkage between social research and public policy. In: LYNN, L. E. (Ed) **Knowledge and policy:** the uncertain connection. Washington, DC: National Research Council, 1978.

WEISS, C. H. Knowledge creep and decision accretion. **Knowledge: creation, diffusion, utilization**, v. 1, n. 3, p. 381-404, 1980.

WEISS, C. H. Evaluation for decisions: is anybody there? Does anybody care? **Evaluation Practice**, v. 9, n. 1, p. 5-19, 1988.

WEISS, C. H. Have we learned anything new about the use of evaluation? **American Journal of Evaluation**, Champaign, v. 19, n. 1, p. 21-33, mar. 1998a.

WEISS, C. H. **Evaluation**. Upper Saddle River: Prentice-Hall, 1998.

WEISS, C. H.; BUCUVALAS, M. J. The challenge of social research to decision making. In: WEISS, C. H. (Ed) **Using social science research in public policy making**. Lexington, MA: Lexington Books, 1977, p. 213-33.

YEH, J. **A reanalysis of test use data** (CSE Report n. 143). Los Angeles: University of California, Center for the Study of Evaluation, 1980.

YIN, R. K. **Case study research. design and methods**. Thousand Oaks: Sage, 2003.

APÊNDICE

QUESTIONÁRIO DA ENTREVISTA SEMI-ESTRUTURADA I

1. Identificação do entrevistado

a) Comentar qual a participação do entrevistado na avaliação do projeto, qual o seu papel na organização.

2. Ideia da avaliação

a) Como foi introduzida a ideia de avaliar o projeto?
b) Por que o projeto foi escolhido para a avaliação?
c) Quais eram as expectativas em relação à avaliação no período em que se decidiu avaliar?

3. Planejamento e implementação do processo de avaliação

a) Como foi definido o escopo da avaliação e como foi planejada (o que fazer, como, por quem, quando) a realização?
b) Havia um projeto inicial de avaliação ou um acordo formal, especificando as diretrizes gerais do trabalho (cronograma de trabalho, orçamento etc)?
c) Como aconteceu a definição das questões principais, critérios de avaliação e procedimentos para coleta e análise de dados?
d) Como avalia o acesso e a qualidade da informação, dos dados e dos materiais disponibilizados?
e) Como qualifica o processo de colaboração entre os vários parceiros e grupos de interesse durante o processo de avaliação?
f) Qual a frequência da troca de informações e a qualidade da comunicação entre os atores envolvidos durante o processo de avaliação?

4. Fechamento e finalização da avaliação

a) As descobertas da avaliação foram reportadas e disseminadas de maneira apropriada para os usuários interessados?
b) Qual a sua percepção quanto à clareza do(s) relatório(s) de conclusão da avaliação? Como avalia a confiabilidade e consistência da análise?
c) As conclusões e recomendações apresentadas no(s) relatório(s) de avaliação se mostraram úteis ao projeto?
d) Os resultados das avaliações foram condizentes com as crenças e expectativas dos usuários?
e) Ocorreu alguma modificação e/ou melhoria no projeto derivada da avaliação? Poderia citar um exemplo?
f) O relatório de avaliação foi tempestivo? Isso é, ele foi entregue no momento oportuno e previamente planejado?
g) A equipe de avaliadores criou incentivos para facilitar o entendimento das informações contidas nos relatórios, explicando os jargões e vocabulários específicos utilizados pelos avaliadores? Ou seja, a equipe de avaliadores ajudou a equipe do projeto a entender o relatório de avaliação?

5. Percepção geral da experiência da avaliação

a) A avaliação mostrou-se adequada às necessidades do projeto? Dos tomadores de decisão?
b) Como você percebe o clima político na Instituição durante o período de avaliação?
c) Você considera que o projeto avaliado estava *preparado* para receber uma avaliação de impacto desse porte?
d) Qual a sua percepção geral quanto às fraquezas e pontos fortes das avaliações? Quais são eles?