

FELIPE MICHEL SANTOS ARAÚJO BRAGA

Padrões para avaliações de programas educacionais:
um estudo de caso para a avaliação do Programa Poupança Jovem.

Belo Horizonte

2010

FELIPE MICHEL SANTOS ARAÚJO BRAGA

Padrões para avaliações de programas educacionais:
um estudo de caso para a avaliação do Programa Poupança Jovem.

Monografia apresentada à Escola de Governo Professor Paulo Neves de Carvalho da Fundação João Pinheiro como requisito parcial para a obtenção de título de bacharel no Curso Superior de Administração Pública.

Orientador: Bruno Lazzarotti Diniz Costa.

Belo Horizonte

2010

Braga, Felipe Michel Santos Araujo

B813p

Padrões para avaliações de programas educacionais: um estudo de caso para a avaliação do Programa Poupança Jovem./ Felipe Michel Santos Araujo Braga – Belo Horizonte, 2010.

59 f.: il.

Monografia (Graduação em Administração Pública) – Escola de Governo Professor Paulo Neves de Carvalho, Fundação João Pinheiro.

Orientador: Bruno Lazzarotti Diniz Costa

1. Avaliação de Programas Governamentais. 2. Estado para Resultados. 3. Padrões. Programa Poupança Jovem. I Bruno Lazzarotti Diniz Costa. II. Título.

CDU 304 (815.1)

FOLHA DE APROVAÇÃO

Para minha família e para todos aqueles que acreditam que a educação é a chave para tornar Minas Gerais e o Brasil os melhores lugares nos quais se viver.

AGRADECIMENTOS

Ao professor Bruno Lazzarotti pela orientação na realização deste trabalho, pela paciência com meus atrasos e questionamentos, pela indicação de boas leituras e bom exemplo de profissionalismo.

Aos demais professores do curso de Administração Pública, co-responsáveis pela minha formação e graduação, bem como pela de tantos outros csapianos que se dedicam ao serviço público. Agradeço também aos funcionários da Escola de Governo, da coordenação do curso e da biblioteca em especial, sempre atenciosos e prestativos.

Aos colegas e amigos da Fundação João Pinheiro, com os quais trabalhei, aprendi e me diverti, tenha sido no CEPE, no DA CSAP, na Consultoria JP Júnior, nos pátios ou nos corredores.

À inesquecível turma do XVIII CSAP, pelo conhecimento construído junto, pelas risadas de cada dia, pelas festas, pelos bons e inúmeros momentos vividos. Ao IPM, a matilha prateada, equipe que um dia ainda se consagrará campeã no CSAPÃO...

Aos colegas do Núcleo de Estudos em Segurança Pública da FJP, Andreia Santos, Eduardo Batitucci, Elaine, Joyce, Roberta, e do Programa Estado para Resultados da SEPLAG, espaços nos quais fiz meus estágios e percebi mais do que o aperfeiçoamento profissional. À Unidade de Indicadores do EpR a devida reverência pelo primor e pelo carinho de todos os seus membros, Caio, Éber, Eder, Elias, Florence, Mônica, Priscila, Renato e Vanda.

Aos meus pais, irmãos, demais familiares e amigos, todos eles com importante contribuição para minha formação acadêmica e humana. Em especial à Luíza, pelo amor e presença.

A avaliação nasceu no sétimo dia da Criação, quando Deus viu tudo o que Ele havia criado e disse: "É bom." Com apenas essa frase Deus torna-se o primeiro avaliador do mundo. No entanto, para alguns, a avaliação nascida no céu resultava insuficiente e por demais subjetiva. Portanto, ao ouvir a avaliação de Deus (auto-avaliação, por sinal), o diabo se opôs e insistiu: "Como Você sabe que é bom? Como mediu? Com que indicador julgou a bondade da Sua criação? Com o que a comparou?" E assim nasceu a avaliação, tal como a conhecemos hoje, no fogo do inferno!¹

(MOKATE, 2002).

¹ Tradução própria. Original em espanhol.

RESUMO

O trabalho parte do estudo de algumas dimensões de desempenho, como eficácia, eficiência e *accountability*. A problematização da situação do ensino em Minas Gerais, principalmente o de nível médio, subsidia o debate sobre a importância das políticas educacionais e da avaliação dos programas públicos a elas vinculados.

Este trabalho utiliza o modelo desenvolvido pelo *Joint Committee on Standards for Education Evaluations* para realizar um estudo de caso sobre a avaliação do Programa Poupança Jovem, a “PESQUISA JOVEM: avaliação do ensino médio”. A avaliação do Programa Poupança Jovem foi elaborada e levada a campo pelo CEDEPLAR/UFMG e contratada pelo Governo de Minas Gerais, representado pela equipe do Estado para Resultados. Esse núcleo sumariza as características da gestão por resultados implementada pelo governo de Minas Gerais, marcadamente a partir de 2007, e é tomado como referência da continuação das políticas que compuseram o chamado choque de gestão.

O modelo analisado é composto por quatro grupos temáticos: Utilidade, Viabilidade, Adequação Ética e Rigor ou Exatidão. Cada um desses grupos é composto por um conjunto de padrões cujo atendimento é considerado adequado para a boa prática avaliativa de programas públicos. Já o estudo de caso da avaliação do Programa Poupança Jovem demonstrou que ela é muito bem qualificada e pode servir de *benchmark* para outras avaliações.

Palavras-chave: Avaliação de Programas Governamentais. Estado para Resultados. Padrões. Programa Poupança Jovem.

ABSTRACT

This monograph commences with an analysis of some dimensions of performances such as efficiency, effectiveness and accountability. The educational circumstances in Minas Gerais, especially at the high school level, subsidizes the debate concerning the importance of educational policies and evaluations of public programs related to these policies.

This work uses the model developed by the Joint Committee on Standards for Education Evaluations to conduct a case study on the evaluation of the Youth Savings Program, called "PESQUISA JOVEM: avaliação do ensino médio." The Youth Savings Program evaluation was developed and applied by CEDEPLAR/UFMG and contracted by the Minas Gerais Government, represented by the State for Results team. This organization summarizes the main characteristics of results management model implemented by the government of Minas Gerais, since 2007, and it is taken as a policy continuation that became a reference of the so-called shock management.

The model analyzed has four thematic groups: Utility, Feasibility, Propriety and Accuracy. Each one consists of a set of standards taken as appropriate for the evaluation practice of public programs. The Youth Savings Program evaluation has been accepted as a very well-qualified evaluation in the study case and serves as a benchmark for other ones.

Keywords: Program Evaluation. Standards. State for Results. Youth Savings Program.

LISTA DE ILUSTRAÇÕES

Quadro 1: Padrões de Utilidade.....	29
Quadro 2: Padrões de Viabilidade.....	30
Quadro 3: Padrões de Adequação Ética	30
Quadro 4: Padrões de Rigor ou Exatidão	31
Quadro 5: Resumo da avaliação da “PESQUISA JOVEM: avaliação do ensino médio”	53

LISTA DE SIGLAS

CEDEPLAR - Centro de Desenvolvimento e Planejamento Regional de Minas Gerais

CSPRO - *Census and Survey Processing System*

DeGEval - *Evaluation Society of Austria and Germany* –

DHS - *Demography and Health Survey*

EPR – Estado para Resultados

FJP – Fundação João Pinheiro

GPR – Gestão por Resultados

IDH - Índice de Desenvolvimento Humano

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais

JCSEE - *Joint Committee on Standards for Education Evaluations*

LDO - Lei de Diretrizes Orçamentárias

LOA - Lei Orçamentária Anual

MEC - Ministério da Educação

NGP – Nova Gestão Pública

PMDI - Plano Mineiro de Desenvolvimento Integrado

PNAD - Pesquisa Nacional por Amostra de Domicílios

PJ – Poupança Jovem

PPAG - Plano Plurianual de Ação Governamental

SEVAL - *Swiss Evaluation Society*

UFMG – Universidade Federal de Minas Gerais

SUMÁRIO

1. Introdução.....	1
2. Referencial teórico.....	6
2.1 Dimensões do desempenho: <i>accountability</i> , eficácia e eficiência.....	6
2.2 A importância de avaliar as políticas, especialmente as educacionais	8
2.3 A avaliação como premissa da gestão por resultados.....	13
3. O modelo de gestão por resultados em Minas Gerais	20
4. Um modelo com diretrizes para avaliações de qualidade.....	25
4.1 Surgimento do <i>Joint Committee on Standards for Education Evaluations</i> e disseminação no mundo.....	25
4.2 Padrões da JCSEE: uma visão geral	28
4.2.1 Padrões de Utilidade	28
4.2.2 Padrões de Viabilidade	29
4.2.3 Padrões de Adequação Ética.....	30
4.2.4 Padrões de Rigor ou Exatidão.....	31
5. Estudo de caso: a avaliação do programa Poupança Jovem.....	33
5.1 A importância da educação como foco de políticas públicas	33
5.2 O Programa Poupança Jovem.....	37
5.3 Aplicação do modelo da JCSEE à avaliação do Programa Poupança Jovem.....	39
5.3.1 Padrões de Utilidade	40
5.3.2 Padrões de Viabilidade	43
5.3.3 Padrões de Adequação Ética.....	44
5.3.4 Padrões de Rigor ou Exatidão.....	46
6. Considerações finais	50
APÊNDICE A – Questionário aplicado na entrevista.....	59

1. Introdução

A gestão por resultados (GpR) é uma das componentes de um movimento heterogêneo que, destacadamente no final do século XX e início do seguinte, orientou parte significativa das reformas do Estado no mundo. A esse movimento, composto por múltiplas facetas, muitas vezes se deu o nome de Nova Gestão Pública (NGP).

Parte relevante dessa componente se pauta pelo processo de avaliação das políticas públicas e disseminação dos seus resultados, considerados pilares fundamentais para o desenvolvimento da democracia. A hipótese central é que a orientação gerencial no setor público – guiada por objetivos e para resultados - conduziria a um melhor custo-benefício das decisões governamentais, sem, no entanto, gerar efeitos (suficientemente) negativos sob outros objetivos estatais que justificassem abandonar esse norte. Assim sendo, pelo menos no caso das avaliações, seus resultados poderiam aprimorar o programa político avaliado, na medida em que descobrissem ou facilitassem as soluções para os problemas encontrados.

A gestão por resultados supõe que o Estado entregue os serviços e produtos exigidos pela sociedade, bem como que, caso não entregue ou não produza como solicitado, dê publicidade aos motivos que levaram a ser assim. Daí a importância da avaliação, pois ela forma juízo sobre os programas governamentais analisados, examina aqueles que melhor aproveitam os escassos recursos públicos para gerar impacto, para promover o bem-estar. Portanto, responde a apelos da *accountability*.

No Brasil o aumento das experiências de avaliação foi concomitante à dispersão da discussão sobre a reforma do Estado e a busca por sua modernização. Entretanto, “[a] experiência brasileira em avaliação de políticas e programas públicos é anterior ao movimento de modernização da gestão pública” (NADER, 2005). Já na década de 90, segundo Iaies (2003), quase todos os países da América Latina criaram sistemas de avaliação da qualidade. Mas não necessariamente esses sistemas repercutiram mudanças nos programas avaliados, que os levassem a aperfeiçoamentos, tampouco seguiram padrões de funcionamento ou metas de alcance.

Estabelecer padrões e diretrizes² para as avaliações é oportuno. Na ausência de claras definições sobre o que determina se uma avaliação é boa ou ruim, de diretrizes que mensuram

² Diretrizes, padrões, guias, princípios, estandartes são todos termos utilizados arbitrariamente na literatura sobre o tema, sempre se referindo às normas e regras citadas nesse texto. As fontes em inglês, contudo, referem-se sobremaneira a *standards*. Analogamente, as fontes em espanhol referem-se a *estándares*.

a qualidade da avaliação, pode-se perder o foco e produzir avaliações que não atendem a todos os padrões desejados. O processo de avaliação, por si mesmo, refere-se à determinação do mérito ou valor e, portanto, é um processo que envolve a identificação e o acompanhamento de algumas normas. A avaliação de um programa educacional, por exemplo, é uma pesquisa, com base nos padrões observados, sobre o desempenho do que é avaliado. Logo, determinar as características de um modelo que sirva como guia metodológico, de modo a tornar possível o pareamento das avaliações sobre as políticas educacionais, ainda que de programas relativamente distintos, pode conduzir na direção desejada, de consolidar a excelência como referência. Permitir o pareamento das avaliações, torná-las comparáveis, é uma forma de incentivar, entre elas e as políticas públicas que avaliam, o intercâmbio de acertos e de meios exitosos para contornar problemas. É uma tentativa de estimular a aprendizagem organizacional, de evitar que seja como aponta Mokate (2002):

Não é exagero sugerir que a avaliação tem sido percebida nos setores sociais como um diabo ou como um "monstro" – um dever gigantesco imposto aos gestores e executores sem que eles sintam nenhum amor pelo processo nem percebam que tenha alguma utilidade direta aos seus processos gerenciais ou decisórios. A avaliação como um "monstro" tem sido suportada como uma tarefa que tem que ser feita, por ter sido imposta por uma exigência legislativa, por tecnocratas do governo central ou por agências financiadoras. Aqueles que enxergam traços de "monstro" no processo de avaliação sofreram, tipicamente, avaliações sofisticadas, cujos resultados estão traduzidos em grandes relatórios, que decoram a biblioteca de um gerente ou de um tecnocrata, sem maior contribuição para o desenho ou a gestão da iniciativa avaliada.³

As avaliações não deveriam apenas se tornar frequentes, deveriam gerar aporte de valor. A própria Mokate (2002) concluirá que as avaliações são válidas e devem ser aprimoradas. Para ela “no lugar de se esperar que sejam produzidas mudanças gerenciais profundas para então avançar sobre os processos avaliativos, poderíamos utilizar os mesmos processos avaliativos para promover profundas mudanças gerenciais”⁴.

Parece ser esse o objetivo do aparato estatal desenvolvido em Minas Gerais, especificamente a partir de 2003, com a implantação do choque de gestão⁵ e, principalmente, após a adoção do modelo de gestão por resultados. Utilizar de processos avaliativos para promover profundas mudanças gerenciais. Outrossim, é tendência em outras regiões, como nos países membros da OCDE, onde

³ Tradução própria. Original em espanhol.

⁴ Tradução própria. Original em espanhol.

⁵ Sobre o choque de gestão em Minas Gerais ver Vilhena *et al.* (2006).

[há] indícios de que o interesse na avaliação como ferramenta de gerenciamento esteja aumentando. Em relação a esforços anteriores de avaliação, há atualmente expectativas mais realistas e uma aceitação mais ampla de metodologias menos rigorosas, que levam a um maior entendimento quanto ao uso dessa ferramenta por parte das organizações. (ALA-HARJA; HELGASON, 2000).

Porém, é específica a conjuntura de Minas Gerais e o legado das avaliações existentes. Mostra-se interessante a tarefa de avaliar as avaliações disponíveis, aprender com elas para melhor executar as potenciais avaliações por vir.

Nesse cenário, esse texto foca-se em discutir parâmetros ideais para as boas avaliações de programas educacionais. Para tanto, é feita uma revisão bibliográfica acerca da gestão por resultados e das orientações para avaliações permanentes que dela se derivam, não pontuais ou avulsas. O foco em resultado conjectura que sejam elaborados trabalhos capazes de indicar quão acertados são os programas educacionais em funcionamento em Minas Gerais, se alcançam seu público-alvo e se produzem as mudanças previstas. Além disso, a avaliação tem a potencialidade de retroalimentar o programa, isto é, de identificar problemas e, se possível, propor alternativas para solucioná-los, bem como fazer com que essas propostas cheguem aos tomadores de decisão. Importante destacar: os avaliadores nem sempre possuem a autoridade sobre o programa avaliado que lhes permitiria impor sanções ou recompensas conforme suas descobertas sejam ou não seguidas e instauradas. Essa tarefa cabe mais aos formuladores e gestores, ou seja, aos políticos e aos membros das equipes que lideram os projetos.

Posteriormente, discute-se o modelo de diretrizes recomendadas pelo *Joint Committee on Standards for Education Evaluations* - JCSEE⁶, que servirá como base para o julgamento qualitativo da avaliação do programa Poupança Jovem, em curso em Minas Gerais. O longo e sistemático trabalho desenvolvido pelo JCSEE para estabelecer os parâmetros que guiam as avaliações educacionais nos Estados Unidos e no Canadá, reforçado pela replicação do modelo, com adaptações, por instituições de países como França, Alemanha, Inglaterra, Suíça e outros, transmite a importância dos padrões discutidos. Cabe lembrar a definição do próprio Comitê (JCSEE, 1994): “[um] *standard* é um princípio mutuamente acordado pelas pessoas envolvidas na prática profissional, que, se cumprido, irá melhorar a qualidade e a justiça daquela prática profissional, por exemplo, a avaliação”⁷. O caminho para a profissionalização da prática avaliativa é seguir os padrões ou princípios mutuamente acordados pelas pessoas engajadas nessa prática profissional. A maturidade do modelo do

⁶ Tradução própria: Comitê Conjunto para Diretrizes da Avaliação Educacional. Ao longo do texto, contudo, a sigla em inglês, JCSEE, será mantida como referência.

⁷ Tradução própria. Original em inglês.

JCSEE, decorrente da qualificação e experiência dos profissionais que o compõem, lapidado pelo tempo decorrido desde o primeiro lançamento, em 1981, justifica a sua escolha para o presente estudo.

Essas diretrizes servem para programas educacionais muito diferentes, com atributos e consequências pretendidas bastante distintas, por isso, não é trivial estabelecer parâmetros comuns. No entanto, dada a revisão bibliográfica anterior e o estudo pormenorizado feito pelo Comitê, descrito ao longo do texto, puderam-se levantar quais são os atributos comuns e desejáveis das avaliações. Eles elevam a eficácia, eficiência, equidade e sustentabilidade dos programas⁸.

Já o estudo de caso, sobre a avaliação do programa Poupança Jovem, foi delimitado pela disponibilidade do mesmo no âmbito do programa Estado para Resultados (EpR). A avaliação do programa Poupança Jovem está em andamento e é a mais avançada entre as avaliações contratadas, pelo programa Estado para Resultados⁹, para os programas da Secretaria Estadual de Educação de Minas Gerais. Para cumprir a tarefa de acompanhar e avaliar os resultados das políticas públicas estaduais a equipe do Estado para Resultados recorre ao apoio de outras instituições, como a Fundação João Pinheiro (FJP) e o Centro de Desenvolvimento e Planejamento Regional de Minas Gerais (CEDEPLAR), que possuem, entre seus servidores, especialistas em avaliação de políticas públicas. Os produtos dessas avaliações, contudo, não são criticados por um modelo preconcebido. Os padrões que mensuram a qualidade das avaliações não estão sistematicamente organizados e discutidos. Todavia, o modelo desenvolvido pelo JCSEE, localizado na Universidade de Iowa, nos Estados Unidos, e adotado por organizações de peso em outras regiões do mundo, mostra-se como útil ferramenta para incorporação de padrões e pode ser adotado como metodologia explícita de avaliação.

O estudo de caso apontou que a avaliação do Programa Poupança Jovem foi satisfatória na grande maioria dos quesitos, servindo de *benchmark* para as avaliações por vir. No entanto, recomendações quanto à maior divulgação dos resultados para outros grupos de interesse além dos contratantes da avaliação e maior transparência no dispêndio de recursos são

⁸ Karen Mokate (2001) publicou um artigo que discute amplamente esses e outros termos (eficácia, eficiência, equidade e sustentabilidade).

⁹ O Programa Estado para Resultados (EpR) foi criado pela Lei Delegada nº. 112, de 25 de janeiro de 2007, com os objetivos declarados, entre outros, de acompanhar e avaliar os resultados das políticas públicas estaduais e oferecer conhecimento público das metas e resultados relacionados à gestão estratégica do governo, de forma a contribuir para o seu controle social. Além disso, o programa apresenta uma estrutura provisória, extinguindo-se em 31 de janeiro de 2011.

realçadas. Observou-se também a necessidade, por parte do governo, de receber os relatórios dentro dos prazos acordados, de modo a facilitar a realimentação do programa com os resultados obtidos em sua avaliação, via relatórios parciais tempestivos.

O estudo foi baseado em pesquisa bibliográfica. O modelo do JCSEE fundamentou o estudo de caso, elaborado a partir de dados secundários, obtidos do contrato de avaliação, de relatórios parciais entregues pelo CEDEPLAR ao Estado para Resultados, de informações no site do Programa Poupança Jovem¹⁰ e de uma entrevista com o Empreendedor Público Eder Sá Alves Campos, responsável por conduzir a avaliação em nome do Programa Estado para Resultados. A entrevista foi informal, não estruturada, com o objetivo de obter uma visão geral do problema pesquisado. O informante é um elemento chave, visto que é o responsável, no Governo, pelo contato com os pesquisadores e pelo acompanhamento da avaliação propriamente dita.

O presente estudo é formado por essa introdução, seguido de um capítulo com o referencial teórico. No capítulo três é apresentado o modelo de gestão por resultados em Minas Gerais. O capítulo quatro apresenta o modelo do JCSEE e sua disseminação no mundo. O capítulo cinco traz o estudo de caso sobre a avaliação do Programa Poupança Jovem e o último capítulo explora algumas considerações finais.

¹⁰ O site do programa é: <http://www.poupancajovem.mg.gov.br>.

2. Referencial teórico

Para tornar inteligível a organização do referencial teórico optou-se por discriminar separadamente pontos cruciais para o entendimento do problema colocado como cerne desse estudo. Discute-se, separadamente, algumas dimensões do desempenho das políticas públicas, a importância de avaliar as políticas públicas e o estado da arte da avaliação de políticas públicas.

2.1 Dimensões do desempenho: *accountability*, eficácia e eficiência

A discussão sobre o desempenho governamental é complexa, pois não há consenso sobre o que determina se um governo tem alta performance ou não. Para alguns, os critérios para essa decisão devem se pautar pelo alcance das políticas promovidas, quando eficazes e eficientes. Para outros, a aprovação popular é o requisito indispensável para julgamento do desempenho governamental. Um bom governo será aquele que tiver altas taxas de aprovação popular e for, portanto, considerado exitoso pelos cidadãos. Há também aqueles que se preocupam com o rendimento democrático dos governos: sua capacidade de responder às demandas da sociedade e tratá-las com transparência.

As avaliações das políticas públicas também são pautadas por um conjunto de critérios, ou dimensões de desempenho, entre os quais se destacam a *accountability*, a eficácia e a eficiência. Essas dimensões, se não alcançadas, determinam o desempenho sub-ótimo da política ou do programa governamental em relação aos resultados. Não são absolutas, porque outras dimensões também são relevantes, como equidade e sustentabilidade. Essas, no entanto, estão mais relacionadas ao desenho e aos processos dos programas, enquanto aquelas são largamente usadas quando se fala em resultados de políticas públicas, pois procuram traduzir o desenvolvimento e a execução da política pública como um todo, embora não haja consenso sobre o que cada termo comunica. Uma breve discussão sobre os conceitos de *accountability*, eficácia e eficiência não poderá esgotar o tema, controverso na literatura, mas poderá expor a problemática dessas dimensões no ambiente das políticas públicas.

O termo *accountability* não tem tradução direta para o português, sendo, via de regra, associado às idéias de transparência e responsividade ou responsabilização dos governantes.

Do Glossário de Gestão Pública da ENAP (2007) vem a seguinte definição:

Accountability: Uma das finalidades da avaliação de políticas públicas, significando estabelecer elementos para julgar e aprovar decisões, ações e seus resultados. Facilita a responsabilização do gestor/agente tanto pela tomada de

decisões como por sua implementação, obrigando-o a prestar contas de suas ações e também de possíveis omissões.

A idéia de responsabilização transcende a idéia da simples prestação de contas. Não se restringe, pois, à necessidade de justificar e legitimar a discricionariedade daqueles que exercem o poder público. Deve incluir também a possibilidade de sanção, se observado mau comportamento. Para discriminar as sanções há que se destacar a diferenciação de Guillermo O'Donnell (1999) entre *accountability* vertical e horizontal. A primeira refere-se às atividades de fiscalização e controle ascendente dos cidadãos e da sociedade civil sobre governantes e burocratas, como o processo eleitoral, plebiscitos e ações da mídia que busquem expor transgressões e delitos de agentes públicos no exercício do poder. A segunda funciona por intermédio de instrumentos institucionalizados de controle e fiscalização mútua, na forma de freios e contrapesos, entre os poderes executivo, judiciário e legislativo.

Mas a interpretação mais relevante para esse trabalho é da *accountability* entendida como forma de controle de resultados da administração pública. Trata-se de responsabilizar o poder público pelo desempenho das políticas, programas e projetos governamentais. Nesse sentido, a avaliação em si, pautada por diretrizes, pode trazer benefícios significativos para a sociedade. Nas palavras de Ceneviva (2005):

Ademais, ponto que merece consideração, os benefícios dessa forma de controle, conforme sugerido pela literatura, parecem não se limitar às sanções que os cidadãos podem impor aos ocupantes de cargos públicos; mas, sobretudo, ao fato de que tais instrumentos de fiscalização orientam-se por diretrizes que têm o potencial de melhorar o debate público e, nesse sentido, prestam-se não só como forma de controle, mas igualmente como aprendizado cívico.

Já a eficácia é tida como o “resultado obtido da comparação entre as realizações e os resultados reais com os que foram estabelecidos” (ENAP, 2007). Para Karen Mokate (2001) uma intervenção é eficaz se conquista o que devia conquistar, no tempo previsto e com a qualidade esperada: uma organização será eficaz se cumprir cabalmente a sua missão, enquanto um programa será eficaz se conquistar os objetivos que lhe deram razão de existir. Ou seja, a eficácia de um programa governamental depende dele produzir os efeitos desejados sobre a população beneficiária e os impactos almejados sobre o conjunto da população e do meio ambiente. Nesse contexto, eficácia é sinônimo de efetividade, embora não haja consenso na literatura sobre isso, com outros autores diferenciando a efetividade por tê-la como a comparação entre os resultados (previstos e não previstos) e os objetivos.

Por outro lado, a eficiência associa-se à otimização de recursos, “é a obtenção do produto esperado com o menor custo possível. Em outras palavras, é a capacidade de produzir os

resultados esperados/desejados com o menor dispêndio de recursos (humanos, materiais e financeiros)” (ENAP, 2007).

Uma iniciativa não pode ser eficiente sem ser eficaz. Para ser eficiente é necessário cumprir integralmente os objetivos, logo, ser eficaz, e, além disso, fazê-lo sem desperdiçar recursos ou insumos. A preocupação com a eficiência nos dá duas alternativas: para determinados objetivos, cumpri-los integralmente ao mínimo custo ou, para determinado custo, gerar o maior ganho.

Resumidamente, conforme Cavalcanti (2006), a eficiência dos programas políticos é alcançada conforme são respondidas as necessidades elencadas, de acordo com os recursos utilizados e o desenvolvimento da gestão. Se não há esforços duplicados ou sobreposição de políticas, então há eficiência, não somente no programa em si, mas nas ações governamentais como um todo.

2.2 A importância de avaliar as políticas, especialmente as educacionais

Para avançar nesse trabalho cumpre discutir o que se define como avaliações de políticas públicas, qual seu escopo, sua validade, seu propósito. Cabe também observar com mais detalhes o que é adotado como impacto e porque sua avaliação é relevante.

É comum encontrar a definição de economia como a ciência que estuda o emprego dos recursos escassos na produção de bens alternativos. Há sempre um custo de oportunidade ao optar por determinado produto ou serviço, dados todos os demais produtos ou serviços que, pela própria escassez dos recursos, deixarão de ser produzidos. Na discussão sobre políticas públicas encontra-se o mesmo problema. Enfrentar a intensa demanda por serviços sociais com recursos bastante limitados é especialmente crítico em países em desenvolvimento, nos quais cada gasto deve objetivar a maximização do impacto na redução da pobreza. (BAKER, 2000).

A solução, via discussão, planejamento e implementação de uma política pública, não é independentemente alcançada. Uma série de circunstâncias pode permitir ou impedir o sucesso no alcance dos resultados previstos. Via de regra:

[...] os programas públicos não possuem um regulador automático que mostre se as atividades efetuadas podem ser realizadas de forma mais eficiente, em que momento deixaram de ser produtivas, ou em que circunstâncias os recursos utilizados podem ser deslocados para outras atividades do mesmo tipo, produzindo um efeito mais significativo na população. (SULBRANDT, 1994).

As avaliações de políticas públicas, inclusive as educacionais, têm potencial para contribuir para a qualificação dos programas governamentais da área, embora não necessariamente o façam. Equivocadamente, as avaliações podem ser vistas como leituras definitivas da realidade, quando devem, oportunamente, influenciar o sistema ou o programa avaliado, como leituras parciais, sucessivas, melhorando-os. Algumas das preocupações principais para aumentar a governança, de modo a criar valor público, são aproximar Estado e sociedade na gestão da *res publica*, aumentar a *accountability* e a efetividade das políticas públicas. O desenho institucional ideal, baseado na transparência e responsabilização, pode ser alcançado na medida em que

[...] a tarefa de reformar o Estado consiste, por um lado, em equipá-lo com instrumentos para uma intervenção efetiva e, por outro, em criar incentivos para que os funcionários públicos atuem de modo a satisfazer o interesse público. Alguns desses incentivos podem ser gerados pela organização interna do governo, mas não bastam. Para que o governo tenha um desempenho satisfatório, a burocracia precisa ser efetivamente supervisionada pelos políticos eleitos, que, por sua vez, devem prestar contas aos cidadãos. (PRZEWORSKI, 1998).

A avaliação pode apoiar a tomada de decisões, para se alcançarem as metas estabelecidas e para melhorar o desempenho da administração pública. Por outro lado, podem ser avaliados os resultados obtidos (esperados e não-esperados), com a finalidade de mensurar o desempenho e promover a prestação de contas da burocracia e dos representantes políticos responsáveis pela condução das políticas governamentais.

Em suma, parte da supervisão das políticas públicas, que devem responder com zelo às carências e anseios dos cidadãos, pode ser feita através das avaliações. Para isso, as avaliações devem contribuir para com a complexa engrenagem que move as políticas, alimentando o sistema com informações novas, bem como o reorientando.

Marjukka Ala-Harja e Sigurdur Helgason (2000) adicionam a qualificação do gasto e o aprendizado organizacional aos objetivos da avaliação:

Os principais objetivos das avaliações são melhorar o processo de tomada de decisão, a alocação de recursos e a prestação de contas. Tais metas podem ser alcançadas por meio de informação ao público, aos principais agentes de decisão e do encorajamento de processos já instalados de aprendizado organizacional. As avaliações devem ser parte de um arcabouço mais amplo de gestão de desempenho. Podem complementá-lo ou melhorá-lo, mas nunca substituí-lo.

As avaliações de políticas públicas são interessantes mecanismos potenciais para aumentar os benefícios das próprias políticas públicas ou para o público-alvo. No campo educacional, em especial, considerado como potencial espaço para a diminuição das desigualdades sociais e econômicas, é importante verificar se há construção da equalização das oportunidades, ou viabilização dos requisitos para essa construção. Uma forma de buscar essa construção é por

intermédio das políticas públicas, quando mudanças positivas são respaldadas por resultados favoráveis, tanto de direito como de fato¹¹.

A avaliação da política pública, bem feita, permite averiguar esses efeitos, averiguar o alcance do programa e ajudar na proposta de alternativas para a solução dos problemas observados. O problema de uma incompleta avaliação está associado à inadequada apropriação dos resultados das avaliações pelos beneficiários, planejadores e executores. Além disso, a avaliação não deve ser pontual, nem tomada como um fim em si mesma, mas contínua, ou realizada como sucessivas ponderações, para acompanhar o desenvolvimento da política que critica e promover a melhoria dos resultados conquistados. “A avaliação não visa conhecimento como um fim em si mesmo, mas como um meio para a ação” (ALA-HARJA; HELGASON, 2000).

Baker (2000) aponta que, na literatura sobre o tema, recomenda-se que uma avaliação global inclua o monitoramento, a avaliação de processos, a avaliação de custo-benefício e a avaliação de impacto. O monitoramento deve permitir respostas contínuas acerca do estado de implementação dos programas, identificando problemas específicos na medida em que surgem. O monitoramento é:

[...] um processo sistemático e periódico de análise da gestão, funcionamento e desempenho de programas e projetos. Tem como objetivo identificar desvios na execução das ações, entre o programado e o executado, diagnosticando suas causas e propondo ajustes operacionais, com vistas à adequação entre o plano e sua implementação. (CAVALCANTI, 2006).

A avaliação de processos, ou formativa, deve se preocupar em como o programa funciona e nos problemas de entrega dos serviços e produtos. A avaliação de custo-benefício examina os custos, em particular, como eles se relacionam aos potenciais usos alternativos e aos benefícios que são produzidos por intermédio do programa. E, ponto crucial na ótica da sociedade, as avaliações devem se preocupar com os impactos.

Os resultados das políticas públicas devem gerar impactos. Impacto entendido como a mudança social provocada pela mudança política, pelo programa ou projeto executado. Vale lembrar que a preocupação com o impacto é específica da administração pública. Aos entes privados não importa com tamanha profundidade os resultados da sua atuação. À sociedade,

¹¹ Como resultado de direito tomamos aquele elaborado e desenvolvido por instituições e organizações reconhecidas oficialmente. Por outro lado, quando o resultado é reconhecido pela sociedade, ou pelo mercado, fala-se em resultados de fato. Espera-se que haja correspondência entre esses tipos de resultado, ou seja, que os resultados de fato tenham correlação com os resultados de direito e vice-versa. Na medida em que a população se apropria das avaliações divulgadas para formar a própria opinião, cresce a relevância dos resultados de direito na construção dos resultados de fato. No entanto, se a opinião pública é considerada ao avaliar o resultado do programa governamental, então é possível falar no resultado de fato construindo o resultado de direito.

sim. Isso porque se entende como impacto “a relação causa-efeito entre a produção realizada e a mudança social provocada” (SERRA, 2008). E são as avaliações que assumem a tarefa de medir - ou ao menos tentar medir - como as metas políticas são revertidas em produtos, ou resultados, e como esses produtos geram o impacto social desejado.

Cabe ressaltar que é difícil estabelecer as relações entre o produto e o objetivo de mudança social, mas é possível fazê-lo, considerando a “relação direta entre a dificuldade de avaliação do impacto e a distância causal ou mudança entre o produto e os efeitos que têm sobre a ação social” (SERRA, 2008). Outrossim, é desejável, sob o ponto de vista social, fazê-lo, para cumprir o objetivo de ampliar - e aplicar recursos públicos em - políticas que têm modelos causais explícitos e alcançam êxito em suas metas. Não raro políticas são implementadas tendo como embasamento teórico modelos causais implícitos, nos quais os resultados das ações são presumidos, mas não testados. Testá-los significa avaliar seus processos, para determinar as relações causais a eles inerentes, ou avaliar seu impacto, para determinar o valor público adicionado ao bem-estar dos beneficiários do programa em funcionamento. Trata-se de buscar reforçar a qualidade das políticas públicas através do comprometimento do governo em fazer política baseado em evidência. A evidência pode ser alcançada por intermédio da avaliação, que é “um exame sistemático e rigoroso, a partir de critérios específicos, dos logros obtidos por uma política ou programa social governamental com relação às metas propostas e os efeitos esperados nos grupos sociais beneficiários” (SULBRANDT, 1994).

O exercício de avaliar, contudo, não tem uma forma única, mas pode ser resumido em três tipos: *ex-ante*, concomitante ou *ex-post*. A primeira é de natureza prévia, destinada a avaliar a necessidade de agir e destinada a estabelecer a base para o projeto. Alternativamente, são possíveis avaliações concomitantes, ao longo da implantação do programa, em harmonia com o monitoramento, mencionado anteriormente. Quando se dá depois de implantada a política, é do tipo *ex-post*. Sulbrandt (1994) afirma que são três as formas clássicas de avaliação *ex-post*: (i) de metas, (ii) de impacto e (iii) de processos. Na primeira são estabelecidas metas e a avaliação busca delimitar com que grau de sucesso elas foram alcançadas. O autor aponta algo muito relevante: o que se constata, nesse caso, é que foram observadas mudanças concomitantes ao programa, mas que não necessariamente os resultados são causalmente atribuídos a ele. Na terceira forma clássica busca-se medir a cobertura do programa social, como se dão seus processos internos e como ele alcança seus beneficiários. A segunda forma, a avaliação de impacto

[...] tem como intenção determinar de forma ampla se o programa tem os efeitos desejados nos indivíduos, nos domicílios ou instituições, e se esses efeitos podem ser atribuídos à intervenção do programa. Avaliações de impacto podem também explorar consequências não intencionadas, quer positivas ou negativas, nos beneficiários.¹² (BAKER, 2000).

O objetivo da avaliação de impacto é mensurar os efeitos líquidos¹³ da intervenção. Para realizar esse cálculo muitas vezes se recorre a métodos quase-experimentais ou busca-se apoio em métodos estatísticos como os utilizados nos modelos causais lineares. Para conferir se os resultados podem ser atribuídos ao programa é preciso avaliar suas relações causais e, para isso, são necessárias, pelo menos, uma medida antes e outra depois do programa, além do uso de mecanismos de controle. Cumpridos os requisitos, essas medidas podem ser avaliadas através de programas estatísticos, que levantam a significância das relações, a plausibilidade das hipóteses causais, o grau de ajuste do modelo à realidade.

Explicada a importância da avaliação das políticas públicas, cabe destacar o prestígio daquelas atribuídas à seara educacional. Como dito, a educação pública é um dos meios factíveis de prover acesso aos ativos que podem libertar as pessoas do ciclo de pobreza (KERSTENETZKY, 2002). Mas a educação pública tem encontrado o que procura? Os programas que a compõem alcançam suas metas? Os cidadãos percebem esses ganhos e refazem suas expectativas em relação aos retornos esperados por mais educação? Esses são alguns dos problemas que são discutidos pelas boas avaliações.

Avaliar e promover melhorias sucessivas nas políticas educacionais, portanto, também é questão de direito e obrigação do Estado. Conforme Marshall (1967) é direito do cidadão adulto o acesso, quando criança, à educação. E educação de qualidade, mais do que em quantidade, pois “é evidente que a maior aprendizagem não depende do aumento do tempo de permanência na escola, mas sim do emprego mais eficaz do tempo” (BRASIL, 2004). A qualidade da educação é medida nas avaliações da proficiência e das competências dos alunos, que dependem do sucesso no desenho de políticas públicas que promovem essas habilidades. As avaliações das políticas educacionais, portanto, cumprem importante papel se forem capazes de contribuir com o sistema. Dessas avaliações espera-se, ainda, que surjam incentivos para maior qualificação das formulações e implementações das políticas educacionais, pela cobrança de eficácia e eficiência dos entes públicos, promovendo um ciclo virtuoso.

¹² Tradução própria. Original em inglês.

¹³ Os efeitos líquidos são “atribuíveis ao programa, livres de possíveis efeitos de outros elementos ‘contaminadores’ presentes na situação” (SULBRANDT, 1994). Trata-se de diferenciar quais resultados são causalmente atribuídos ao programa político.

Ademais, uma avaliação não será isenta de subjetividade. Atribuir valor requer considerar alternativas, conceber opiniões. Mas é importante, para promover a transparência, que as suposições feitas sejam claras e compartilhadas. Desse modo, aqueles que recebem a avaliação ou a acompanham podem, por si mesmos, decidir sobre a significância das avaliações e a relevância de suas descobertas. Daí a importância de um modelo padronizado para o acompanhamento e avaliação dos programas educacionais implementados em Minas Gerais, que tenha para com eles o devido cuidado. A questão que se coloca é que as políticas públicas educacionais devem receber avaliações coerentes e rigorosas. Ou, detalhadamente: que a metodologia empregada deve ser adequada; os resultados devem ser apropriados pelos planejadores, implementadores e beneficiários; o planejamento inicial deve conceber a avaliação, para que ela não esteja atrelada apenas a decisões futuras e independentes do planejamento político original. É possível concluir que são de diversas naturezas as avaliações do sistema educacional que podem ser levadas a cabo no estado de Minas Gerais, de impacto, de processo, de metas. No entanto, há argumentos que fundamentam as decisões sobre qual delas executar, além de estabelecerem quais atributos levam à classificação das avaliações como boas ou ruins. Por isso os estandartes, conforme a orientação da *Swiss Evaluation Society* (SEVAL):

As normas de avaliação da Sociedade Suíça de Avaliação (doravante *Seval Standards*) são destinadas a contribuir para a profissionalização da avaliação na Suíça. Aderir às normas da SEVAL reforçará a credibilidade, qualidade e confiabilidade das avaliações. Avaliações de alta qualidade só podem ser criadas quando todos os envolvidos - isto é, os próprios avaliadores, os patrocinadores das avaliações e outras pessoas que participam na avaliação - trabalharem juntos. As normas da SEVAL, portanto, estabelecem critérios que todas as pessoas envolvidas em uma avaliação têm o dever de defender.¹⁴ (WIDMER, LANDERT e BACHMANN, 2000).

2.3 A avaliação como premissa da gestão por resultados

A história recente da administração pública brasileira é tratada por Abrucio (2007) e por Nader (2005).

A redemocratização, após o governo militar, é tomada como marco referencial para o processo de reforma do Estado. A necessidade de enfrentar a crise do regime autoritário e, sobretudo, do modelo nacional-desenvolvimentista, exigiu soluções modernas. Nesse contexto, a Constituição Federal de 1988 sintetiza as aspirações e orientações alternativas escolhidas, através: da democratização do Estado, favorecida pelo aumento do controle externo; da descentralização política, financeira e administrativa, que permitiu maior

¹⁴ Tradução própria. Original em inglês.

participação dos cidadãos e inovações na gestão pública; da reorganização do serviço público, através da profissionalização da burocracia.

Para Abrucio (2007) “todas essas mudanças trouxeram ganhos à administração pública brasileira; porém, o sentido de cada uma delas não se concretizou completamente por conta de uma série de problemas.” O controle externo não fora estabelecido com autonomia em relação ao executivo, limitando sua atuação. A descentralização falhou porque houve pouca cooperação entre os entes federativos, além da manutenção do patrimonialismo local em regiões do país. O serviço público permaneceu, em parte, corporativista e capenga, com deformidades associadas, na mentalidade nacional, à burocracia: auto-referimento, formalismo exacerbado, hierarquização excessiva, repleto de barreiras para a criatividade e as soluções inovadoras.

Durante o governo Collor a crítica à burocracia foi vinculada à teoria do Estado mínimo e da extinção dos marajás. As medidas de desestatização de Collor estavam diretamente ligadas ao programa de estabilização lançado em seu governo. Mas “[a]s medidas tomadas nesse período foram um desastre. Houve o desmantelamento de diversos setores e políticas públicas, além da redução de atividades estatais essenciais” (ABRUCIO, 2007).

Assim, foi nesse cenário com grande presença do Estado no setor produtivo, crise fiscal, inflação galopante e deterioração do serviço público que se tornou evidente aos olhos do governo a necessidade de uma Reforma do Estado para estabilização e alcance do crescimento sustentado. Foi no primeiro governo de Fernando Henrique Cardoso, em 1995, que foi criado o Ministério da Administração Federal e Reforma do Estado (MARE) com a responsabilidade de idealizar e implementar a Reforma Gerencial a ser aplicada ao Brasil. Vale colocar que, além dos fatores internos, o cenário internacional estava totalmente favorável a esse tipo de política, na medida em que a Europa, principalmente, estava passando pelo mesmo processo.

A Reforma de 1995 partiu de uma série de perguntas de caráter estrutural, a saber:

[...] primeiro, quais são as atividades que o Estado hoje executa que lhe são exclusivas, envolvendo poder de Estado? Segundo, quais as atividades para as quais, embora não exista exclusividade, a Sociedade e o Estado consideram necessário financiar (particularmente serviços sociais e científicos)? Finalmente, quais as atividades empresariais, de produção de bens e serviços para o mercado? (BRESSER-PEREIRA, 2007).

Essa Reforma foi calcada na redefinição dos papéis do Estado, em que este assume mais fortemente a função de regulador e promotor dos serviços sociais, colocando em segundo plano a sua função de executor e prestador de serviços. Essa revisão das funções do Estado

propunha uma profunda reestruturação do aparelho estatal, transferindo para a sociedade e para outras esferas atividades que pudessem ser teoricamente melhor executadas no setor público não-estatal e no âmbito dos governos estaduais e municipais. A reestruturação fora orientada também pela busca de uma adequação entre funções, formas de propriedade e tipos de gestão, de modo que fosse combinada a mudança institucional com a renovação nos métodos de gestão e na cultura das organizações do setor público.

O projeto esteve apoiado em uma idéia mobilizadora: uma administração voltada para resultados, ou seguindo o modelo tido como gerencial. Esse modelo se relaciona ao movimento da Nova Gestão Pública, calcado em desempenho e performance. “É possível resumir a preocupação central da Nova Gestão Pública como a combinação entre a flexibilização da gestão e o aumento da responsabilização da administração pública” (SANO; ABRUCIO, 2008). Costa (2002) resume o que é chamado de Nova Gestão Pública:

Sinteticamente, a NGP consistiria em um núcleo de idéias que enfocam prioritariamente: a qualidade da gestão, a avaliação de desempenho; a desagregação das burocracias em agências que se relacionam em bases contratuais - e se possível monetárias (*user pay basis*) -; o uso de “quase-mercados” e terceirização para estimular a competição; redução de custos e um estilo de gestão que enfatiza metas, contratos periódicos e autonomia gerencial.

Dentro da Nova Gestão Pública, no enfoque da gestão por resultados, destaca-se a discussão da importância da avaliação de políticas públicas. Nesse contexto se concebem alternativas, ligadas à lógica gerencial, à preocupação com resultados, para abordar os desafios inerentes à administração pública. Enfatiza-se, entre outros pontos, que a racionalidade econômica é relevante e que, a partir dela, estabelece-se a procura por maior eficácia e eficiência. Essa procura é pautada por instrumentos gerenciais e democráticos, aos quais se atribui a capacidade de combater os problemas que o Estado enfrenta no mundo contemporâneo. Porém, enquanto alguns defensores do movimento da NGP atacam o formalismo e a rigidez burocrática como defeitos, podem deixar escapar que algumas bases do modelo weberiano podem ser úteis para a modernização, em prol da meritocracia e da separação entre o público e o privado.

Esses instrumentos defendidos pela NGP podem favorecer a atuação do Estado, são de caráter democrático e gerencial, mas não são necessariamente novos. Parte significativa das propostas da NGP pertence a projetos de reforma anteriores, como a avaliação de desempenho, a subcontratação, a privatização, a utilização de técnicas gerenciais privadas na administração pública. Mas a nova gestão pública pode se diferenciar dos movimentos que a antecederam

pelo enfoque dado à articulação dessas práticas, para atingir o núcleo dos problemas fiscais e administrativos das burocracias.

Na burocracia há vícios a serem sanados e virtudes a ampliar. Em ambos os casos instrumentos como a avaliação podem ajudar. Essa articulação, dentro do escopo da lógica gerencial, seria possível se pautada por três propósitos fundamentais, segundo Serra (2008):

- Assegurar a constante **otimização do uso dos recursos** públicos na produção e distribuição de bens públicos como resposta às exigências de mais serviços e menos impostos, maior eficácia, eficiência, equidade e qualidade.
- Assegurar que o processo de produção de bens e serviços públicos (incluindo a concessão, a distribuição e a melhoria da produtividade) seja **transparente, equitativo e controlável**.
- Promover e desenvolver mecanismos internos que **melhorem o desempenho** dos dirigentes e servidores públicos, e, com isso, fomentar a efetividade dos organismos governamentais, visando a concretização dos objetivos anteriores.¹⁵

A democracia e os objetivos descritos acima, que orientam a gestão por resultados e as respostas às demandas cidadãs, formam os principais pilares de legitimidade do Estado atual (SERRA, 2008). Outrossim, as avaliações das políticas públicas, com destaque as políticas educacionais, são orientadas pelo mesmo guia. Pretende-se, com as avaliações, otimizar o uso dos recursos, garantir que aos dispêndios correspondam ganhos progressivamente melhores. Também é finalidade das avaliações tornar o processo de construção política e consecução dos resultados cada vez mais transparente, equitativo e controlável. As avaliações de programas e políticas públicas podem desempenhar um papel importante na concepção ou melhoria de mecanismos de *accountability* na gestão pública (CENEVIVA, 2005). Isso porque o controle de resultados e o controle social são duas formas válidas de participação e controle da sociedade sobre as ações do Estado. Para realizar esse controle torna-se oportuno promover avaliações, diretamente ligadas à questão da eficácia, eficiência e efetividade, da *accountability* e, em um contexto mais amplo, ligadas ao desempenho da gestão pública (WEISS, 1997).

A avaliação de políticas e programas públicos ganha, assim, um lugar de bastante destaque nas propostas de reforma do Estado e constitui-se, de acordo com a NGP, no processo estruturante do controle de resultados. Segundo Mario Quirós (2000) “a responsabilização pelos resultados na administração pública baseia-se na avaliação *ex-post* das políticas e

¹⁵ Grifos no original.

programas, como meio para medir seu desempenho e exercer a prestação de contas dos hierarcas e funcionários públicos”.¹⁶

Do mesmo modo, do escopo avaliativo faz parte a procura por melhorias do desempenho, entendido como o progresso no cumprimento das metas e objetivos estabelecidos.

Por outro lado, é interessante observar que não são únicas a importância e a finalidade das avaliações. A avaliação pode ser usada para diferentes fins em diferentes modelos, inclusive em diferentes contextos de gestão por resultados. O tipo, o conteúdo e o papel desempenhado pela avaliação podem variar de acordo com as suposições e as estratégias de execução política adotadas. Uma distinção típica na literatura é aquela entre a implementação programada e a implementação adaptativa¹⁷. Berman (1978) discute as circunstâncias que podem levar à escolha de cada uma, ou a uma mistura de ambas as estratégias, sendo que cada qual tem uma finalidade para a avaliação.

Os adeptos da implementação programada são aqueles que acreditam que os problemas de execução de um programa político podem ser minimizados, senão eliminados, pela cuidadosa e explícita pré-programação dos procedimentos de implementação. A ênfase está no planejamento. Portanto, as especificações no desenho do programa devem reunir, preliminarmente, clareza, precisão e alta compreensibilidade da política. A decisão política, uma vez tomada, deve ser fielmente seguida.

Por essa linha de raciocínio argumenta-se que eventuais ambiguidades no desenho deixam os implementadores desorientados. Por isso devem ser formulados objetivos específicos, detalhados e presumivelmente consistentes. A ambiguidade poderia se originar também de ter muitos atores com poder de decisão. Assim, é necessário estabelecer uma autoridade final, centralizada, que determina o que e como fazer. Já a potencial resistência à implementação se derivaria da má operacionalização das ações e do programa. Para evitar ou superar a resistência as opções são restringir a liberdade dos executores, estabelecendo o que devem fazer; monitorar o comportamento dos implementadores para que saibam que estão sendo fiscalizados; e usar de incentivos extrínsecos.

Nesse cenário, o objetivo da avaliação é a fiscalização, é medir a correspondência do resultado da política ao pré-especificado, de preferência em relação a metas quantitativas. É

¹⁶ Tradução própria. Original em espanhol.

¹⁷ A implementação programada e a implementação adaptativa são modelos ideais, polares ou antagônicos. Programas políticos são usualmente uma mistura de ambas, ainda que com prevalência de uma delas. Na literatura esses modelos são discutidos também como estratégias *top-down* e *bottom-up*.

tomar a relação entre a decisão política e seu produto como prova que a execução foi realizada conforme foi concebida. Se os resultados forem abaixo das expectativas, a avaliação poderá ser usada de duas maneiras. Primeiro, se verificar que os executores se desviaram do programado, então eles poderiam ser responsabilizados e sanções (ou talvez mudanças na alocação de recursos e incentivos) seriam empregadas. Segundo, se os executores aderiram à orientação e os objetivos não foram alcançados, então as recomendações podem ser modificadas incrementalmente. Incrementalmente, não de forma profunda. A implementação programada assume que a base da decisão política está essencialmente correta. Se evidências avaliativas demonstram o contrário, então a política será julgada falha. Descartada, em vez de revista.

A implementação adaptativa é, no modelo ideal, o contraponto da anterior. Ela é mais pessimista em relação ao planejamento e o concebe passível de cometer falhas e confundir incentivos e respostas esperadas. Para seus adeptos os problemas de implementação surgem da especificação exagerada e rigidez de metas, do erro de não considerar atores relevantes no processo decisório e pelo excesso de controle. A ênfase está na execução, em acompanhar os processos que permitiriam modificações na política, adaptando-a dadas as interações da política com a estrutura na qual ela atua. Os resultados não são automáticos ou assegurados e a realidade parece-se mais com um processo desordenado de aprendizado do que com um processo previsível.

Destarte, opta-se por regras que podem ser revistas. Por conceder arbítrio aos executores do programa - para potencializar o *learning-by-doing*. Por aumentar a participação dos grupos envolvidos - para trazerem mais pontos de vista e enriquecerem o processo.

Na implementação adaptativa a avaliação é tomada como meio de reorientação futura, não como medida da fidelidade a uma decisão inicial e ao plano de operação. É antecipado que o plano original vai sofrer mutações, quando posto em prática, porque ajustes serão causados por condições idiossincráticas de cada local e como desdobramentos do *learning-by-doing*.

Conseqüentemente, a avaliação deverá prover informação para que os ajustes locais e o aprendizado tomem lugar. Os dados recolhidos e encaminhados aos tomadores de decisão serão sobre processos adaptativos e resultados. Se a adaptação não se provar produtiva, haverá ajustes para que possa evoluir. Nesse contexto a formulação não é moldada em concreto, rígida, indiferente às circunstâncias. A implementação adaptativa é, pelo contrário, um meio de procurar especificidade. A avaliação, nesse caso, é usada para reconhecer as

especificidades da política e promover aprendizado. A execução conclui a formulação, criando ciclos de formulação e implementação.

Assim sendo, quer a implementação seja programada, quer adaptativa, percebe-se que a avaliação das políticas públicas é imprescindível para a gestão por resultados, pois concretiza o ideal desse modelo de gestão: pautar a ação governamental pelo valor público entregue à sociedade. Excelentes desenhos de programas políticos, sem a devida execução, podem não ser satisfatórios. Excelentes execuções, mas de políticas mal desenhadas, também não trazem os benefícios desejados. A avaliação pode ser tomada como um dos instrumentos aptos a verificar se os resultados são bons, seja a produção de um bem ou o fornecimento de um serviço público, confirmando a qualidade do desenho e da execução do programa que os prevê. Logo, cabe discutir o papel da avaliação em Minas Gerais e, cumpre, também, discutir como é o modelo de gestão por resultados no estado.

3. O modelo de gestão por resultados em Minas Gerais

A gestão por resultados levada a cabo em Minas Gerais é singular, principalmente por tratar-se de uma adaptação das recomendações da literatura para o contexto político e econômico estadual.

Com base nas premissas de ajuste fiscal e modernização da administração pública o estado de Minas Gerais instituiu, a partir de 2003, o programa Choque de Gestão. Foi adotado o duplo planejamento no governo, associando metas de longo e de curto prazo. O planejamento de longo prazo se materializou no Plano Mineiro de Desenvolvimento Integrado (PMDI), elaborado pelo poder executivo e aprovado pelo legislativo, que continha a visão de futuro “tornar Minas Gerais o melhor estado para se viver”. Além disso, o PMDI abordou os objetivos estratégicos do governo e as áreas de resultado, que são as áreas nas quais se concentrariam os esforços e os recursos, com o objetivo de gerar as transformações sociais desejadas. O PMDI é desdobrado em planejamentos de quatro anos, os Planos Plurianuais de Ação Governamental (PPAG), que trazem os projetos estruturadores, os programas associados, os programas especiais, assim classificados conforme a prioridade a eles atribuída. O PPAG é seguido por outros dois instrumentos instituídos pela Constituição Federal, a Lei de Diretrizes Orçamentárias (LDO) e a Lei Orçamentária Anual (LOA).

Além disso, para que os projetos prioritários pudessem ser executados e os resultados perseguidos, foram estabelecidos novos instrumentos, como os acordos de resultado, bem como novas estruturas e arranjos, a exemplo o escritório de projetos. O acordo de resultado é um contrato celebrado entre o governo, centralizado na Secretaria de Estado de Planejamento e Gestão, e as unidades descentralizadas, com metas e resultados a cumprir, negociados ano a ano. O escritório de projetos, intitulado Gestão Estratégica de Recursos e Ações do Estado (GERAES), monitora intensivamente a carteira de Projetos Estruturadores, participando do estabelecimento dos planos e das metas, bem como acompanhando o desempenho da execução física e financeira. Mas alcançar uma boa prática de gestão pública por resultados não implica somente em lançar mão de mecanismos de avaliação institucional.

A gestão pública por resultados é viabilizada por diversos mecanismos gerenciais, começando pelo planejamento estratégico das ações governamentais - sejam elas organizadas em projetos ou programas; ampliação da flexibilidade gerencial; desenvolvimento de indicadores de desempenho; e, por fim, a avaliação de desempenho, que além da medição das metas estabelecidas, fornece subsídios para retroalimentação de informações para o desenvolvimento de futuras metas gerenciais. (CORREIA, 2007).

O Programa Estado para Resultados é mais do que uma instituição avaliadora. É a referência em relação à segunda geração das políticas do Choque de Gestão mineiro. As mudanças instituídas pelo PMDI 2007-2023, revisão daquele realizado em 2003, reforçaram o modelo focado em resultados, “na medida em que operacionalizaram metas e indicadores de desempenho para o estado” (BECHELAINÉ, 2009). O PMDI 2007-2023 operacionaliza esses objetivos por meio de indicadores, chamados de indicadores finalísticos, que são acompanhados pelo Programa Estado para Resultados¹⁸.

É interessante observar como a implementação de gestão por resultados no estado de Minas Gerais teve sua construção marcada pelo incrementalismo. Na definição de Berman, que contrapõe a implementação programada à adaptativa, percebe-se que o Programa Estado para Resultados nasceu de uma concepção predominantemente programada, mas assumiu progressivamente algumas características de implementação adaptativa.

A instauração do programa decorreu da revisão do planejamento de longo prazo do governo (BECHELAINÉ, 2009). Assim, as razões que o levaram a existir estavam predominantemente ligadas à necessidade do governo de controlar as unidades descentralizadas, de tornar-se capaz de acompanhar e monitorar as políticas que, se implementadas conforme planejadas, levariam, supostamente, aos resultados desejados. A ênfase estava no planejamento, mas migraria para o alcance dos resultados, pela realização de adaptações que foram consideradas necessárias e oportunas para o alinhamento da estratégia à execução.

A reformulação em curso desde 2007 dos Acordos de Resultados propicia ainda mais esse alinhamento, na medida em que pactua com os órgãos da administração pública, os resultados sinalizados no PMDI, a execução dos Projetos Estruturadores, além de itens de melhoria da gestão, racionalização do gasto, e ações estratégicas a serem implementadas. (GUIMARÃES; CAMPOS, 2009).

A princípio, o diagnóstico inicial subsidiaria as políticas a serem implementadas e o sucesso dependeria principalmente de executá-las de acordo com o que fora idealizado. Os incentivos favoráveis às mudanças seriam dados, por exemplo, pelos acordos de resultado e pela remuneração variável vinculada ao alcance das metas pactuadas. Além disso, tomava-se que valores sociais associados às metas, como o consenso sobre a necessidade de queda na taxa de mortalidade ou diminuição do analfabetismo, reforçariam o novo modelo de gestão.

Como explorado por Berman (1978), no entanto, o êxito dos programas políticos não depende exclusivamente do planejamento. O autor destaca que os parâmetros situacionais determinam

¹⁸ O acompanhamento dos indicadores finalísticos é feito pela Unidade de Indicadores. Um estudo de caso sobre o Programa Estado para Resultados, que detalha seu desenvolvimento histórico e funcional, bem como o das unidades que o compõe, foi feito por BECHELAINÉ (2009).

quando há maior chance de sucesso no uso da implementação programada do que na adaptativa. Nesses casos: o escopo de mudança no comportamento dos membros da implementação é mínimo; a validade da tecnologia política é relativamente certa; os membros do sistema de implementação concordam com as metas e meios da política; a estrutura de coordenação da implementação é fortemente acoplada; e o ambiente do sistema de implementação é estável. À medida que se observam diferentes condições nos parâmetros situacionais, diferentes abordagens, vinculadas à implementação adaptativa, tornam-se apropriadas.

O escopo da mudança trata do tipo e da quantidade de esforços que são necessários para fazer com que os membros do sistema de implementação aceitem as mudanças e percebam incentivos para agir em prol das mesmas. Ao Programa Estado para Resultados cumpriu a tarefa de garantir que as mudanças não seriam apenas simbólicas ou aparentes, pois era preciso que as mudanças gerenciais interferissem nos resultados finalísticos medidos pelos indicadores. Para isso, as decisões não poderiam ser arbitrárias, mas envolver os órgãos e equipes executoras, que passaram a negociar os indicadores e os níveis das metas.

De forma a garantir o envolvimento tanto das secretarias de estado quanto de seus servidores, foram oferecidos incentivos interligados aos dois. Aos servidores dos órgãos e entidades que celebraram o acordo de resultados foram concedidos prêmios por produtividade quando conseguiam reduzir as despesas correntes concernentes a suas atividades. Aos órgãos e entidades acordantes foi viabilizada a aplicação dos recursos economizados no desenvolvimento de pessoal, modernização e racionalização dos serviços públicos prestados. Além de garantir maior flexibilidade e autonomia aos órgãos e entidades proponentes, o acordo de resultados ajuda na construção da cadeia causal da promoção dos resultados almejados, uma vez que indicadores intermediários também são mensurados. (CORREIA, 2007).

O segundo parâmetro situacional apontado por Berman é a validade da tecnologia, ou da cadeia causal. A implementação programada tem maior probabilidade de êxito quando as relações causa-efeito estão claras e bem sedimentadas. Destarte, não são geradas críticas ou desconfianças quanto ao potencial alcance dos programas. A mudança estratégica profunda, da implantação da gestão por resultados, diferencia-se nesse ponto, pois parte de um modelo causal implícito, de hipóteses não testadas previamente. Elementos importantes de incerteza surgiram da interação da política com as peculiaridades do sistema de implementação e do seu contexto. Não era suficiente conferir se a execução se dava conforme planejada. A estrutura planejada precisava ser revista e alterada. Nesse caso, a abordagem adaptativa se mostrou oportuna.

O conflito sobre as metas, outro parâmetro situacional, foi articulado pelo Programa Estado para Resultados, que evitou a submissão simbólica dos demais órgãos, ao incluir as secretarias estaduais e suas unidades na definição dos objetivos, incorporando os pontos de vista dessas partes interessadas.

Dado o foco no cumprimento de ações definidas, é necessária maior vinculação entre a alta direção e suas “unidades” operativas, para que sejam definidas as responsabilidades e compromissos mutuamente assumidos. Com o objetivo de fortalecer esse vínculo, acordos que estipulam os resultados a serem alcançados são, geralmente, articulados entre as duas partes. Nesse tipo de acordo, as unidades operativas se comprometem a cumprir com as metas previamente estabelecidas, enquanto a alta direção fornece incentivos que favorecem o alcance das metas. (CORRÊA, 2007).

A estrutura de coordenação trata da arena política, na qual ajustes e barganhas são recorrentes. A aderência a regulações fixas, uniformes, não é a regra, pois as preferências dos agentes envolvidos podem mudar. As características adequadas à implementação adaptativa se manifestam, como a procura por especificidades, pela identificação de como o sistema se diferencia nas unidades operacionais e como as unidades são coordenadas. Nesse sentido, o acordo de resultados prevê a maior autonomia dos órgãos, enquanto o acompanhamento transfere o enfoque para os resultados.

O último parâmetro situacional, a estabilidade do ambiente, considera como as forças ou condições fora do sistema de implementação o afetam. Qual a relevância dos fatores exógenos e a importância dos planos contingenciais. Ambas, em relação à metodologia de implementação da gestão por resultados em Minas Gerais, foram altas. O método de tentativa e erro foi usado, ou seja, “o modelo criado foi se transformando com práticas internalizadas na rotina para adquirir uma estrutura adaptada” (BECHELAINÉ, 2009). O programa não partiu de teorias pré-concebidas, foi construído de forma original e conforme as especificidades que o cercavam. Caracterizou-se pela adaptação ao contexto.

Por tudo isso, a avaliação tem papel híbrido no modelo de gestão para resultados de Minas Gerais. Responde a apelos programáticos e adaptativos. De um lado, ela é tomada como uma forma de controle e fiscalização da burocracia, de prestação de contas, de *accountability*. O diagnóstico foi realizado previamente, foram idealizadas as alternativas políticas para alcançar o impacto desejado e a avaliação seria, então, a responsável por averiguar qual o resultado entregue e a profundidade do mesmo, em relação ao alcance do público-alvo e das mudanças sociais. Seria responsável ainda, conforme ateste ou não a consecução dos resultados almejados, por “justificar a continuidade ou finalização de determinadas políticas públicas” (GUIMARÃES; CAMPOS; 2009). Por outro lado, como o programa Estado para Resultados

tem caráter incremental, de adaptação ao contexto, a avaliação também colabora com a possibilidade de aprendizagem organizacional e reorientação orçamentária.

A ideia de constituir um conjunto de avaliações de impacto remete ao fato da necessidade de ter ao final de um ciclo governamental (no caso em questão, 2007-2010), um conjunto mais confiável de informações acerca da efetividade das principais políticas públicas no Estado.

[Há ainda] a necessidade, provavelmente, de se caminhar para o fortalecimento das avaliações, atrelada às mudanças gerenciais e orçamentárias, que poderiam significar perda relativa de recursos para os órgãos de pior desempenho. (GUIMARÃES; CAMPOS, 2009).

Há a preocupação com a eficácia, com a efetividade. As descobertas das avaliações podem melhorar o processo de tomada de decisão e a alocação de recursos. São, por conseguinte, efetuadas com o propósito de realimentar o processo de formulação das políticas públicas e programas governamentais.

4. Um modelo com diretrizes para avaliações de qualidade

4.1 Surgimento do *Joint Committee on Standards for Education Evaluations* e disseminação no mundo

Um dos objetivos iniciais da pesquisa que conduziu à produção desse texto era estabelecer parâmetros, baseados na revisão bibliográfica anterior, para comparação e ponderação das avaliações disponíveis, de modo a tornar possível o pareamento das avaliações sobre as políticas educacionais, ainda que de programas relativamente distintos.

Elaborar e estabelecer essas diretrizes, no entanto, mostrou-se um exercício impróprio, dado que, no levantamento bibliográfico, percebeu-se a existência de esforços de Organizações Internacionais para criar modelos com escopo muito mais ampliado.

A importância do modelo do *Joint Committee on Standards for Education Evaluations* - JCSEE é descrita por Rodrigues (2009):

Produto de um trabalho extenso, alargado, aturado, moroso, abnegado e rigoroso, que conduziu a um consenso abrangente no seio da comunidade educativa norte-americana, sem dúvida por força da participação empenhada e representativa das grandes organizações científicas e profissionais da educação dos Estados Unidos da América (e, também, do Canadá), os três conjuntos de *standards*, publicados pelo *Joint Committee on Standards for Educational Evaluation* (JCSEE), constituem, por força desse esforço, no seio de uma vasta nação, e pelo lugar que a produção desse país ocupa no plano científico e profissional mundial, uma referência incontornável para todos aqueles cuja actividade laboral envolve directamente a avaliação, sobretudo no âmbito educativo, mas, também, para todos quantos por ela são afectados, o que eleva bastante mais o seu alcance.

O modelo deu relevante contribuição para o cenário das avaliações de políticas educacionais norte-americanas e foi adotado, com adaptações, em países ao redor do mundo. Os *standards* do JCSEE foram extensamente publicados, tendo sido adotados ou adaptados por sociedades de avaliação da Europa, dentre as quais Beywl (2004) destaca: a Sociedade Suíça de Avaliação (fundada em 1996), que aprovou o guia *SEVAL Evaluation Standards* em sua assembleia geral em 2001; a Sociedade Alemã de Avaliação (fundada em 1997), que ratificou a *DeGEval-Standards* também em 2001; a Sociedade Francesa de Avaliação (estabelecida em 1999), que apresentou uma Carta de princípios orientadores para a avaliação de políticas públicas e programas (*Charter of evaluation guiding principles for public policies and programmes*) em 2003; e a Sociedade de Avaliação do Reino Unido (que foi a primeira sociedade europeia, fundada em 1994), que adotou as orientações de boas práticas em matéria de avaliação (*The Guidelines for good practice in evaluation*) em 2003¹⁹.

¹⁹ Tradução própria. Original em inglês.

A Sociedade Suíça de Avaliação (SEVAL) afirma no documento síntese (WIDMER, LANDERT e BACHMANN, 2000) que suas diretrizes para avaliação têm como objetivo contribuir para a profissionalização da avaliação no país. De fato, criar padrões significa tornar replicável um método, significa tomar um referencial e sempre que possível baseado na excelência, de modo a diminuir a variabilidade no uso da técnica. Assim, a perspectiva de futuro da avaliação é ser “uma profissão e uma comunidade, no caminho da estabilização de sua posição nas democracias modernas, embasada em tornar explícitos os padrões para a boa prática avaliativa.”²⁰ (BEYWL, 2004).

Para Rodrigues (2009):

De facto, os *standards* [da JCSEE] (todos e cada um deles) resultam de um rigoroso processo de elaboração e teste, que inclui painéis de formulação, painéis de revisão, testes de terreno, consultas públicas e painéis de validação, e participam de uma filosofia de revisão periódica, que permite acolher os avanços técnicos e científicos constantes e responder aos novos desafios e exigências que se perfilam no terreno da avaliação.

Mas trata-se de um modelo, de um recurso analítico com a intenção de representar a realidade - bastante mais complexa - a partir da definição de um conjunto de algumas variáveis. Os modelos serão sempre limitados, ignorando variáveis potencialmente importantes e com restrições para estabelecer os padrões de comportamento entre as variáveis. Contudo, assumem a tarefa de buscar incluir o essencial e excluir o que não é fundamental.

Cumprir observar, portanto, que mesmo repletas de qualidades, as diretrizes desenvolvidas nesse modelo não são absolutas, ou regras mecânicas, mas princípios orientadores, que não são igualmente importantes ou aplicáveis em todas as situações.

Como benefícios apontados pelos defensores dos padrões de avaliação, Beywl²¹ (2004), co-fundador e membro da diretoria da *Evaluation Society of Austria and Germany* – DeGEval, destaca que:

- Eles promovem o diálogo entre a comunidade de avaliadores e o público em geral, os políticos e administradores sobre o papel potencial da avaliação na democracia e na sociedade;
- Apóiam a criação de uma cultura de avaliação nas administrações públicas e organizações privadas;
- Reforçam a identidade da avaliação como uma profissão, arte e/ou ciência;

²⁰ Tradução própria. Original em inglês.

²¹ Tradução própria. Original em inglês.

- Fornecem evidências sobre o progresso da teoria de avaliação, seus termos e os conceitos fundamentais e, ao mesmo tempo, promovem o desenvolvimento teórico e metodológico;
- Incorporando as principais definições da linguagem de avaliação, tais como valor, interesses, *stakeholders*, audiência, grupo-tratamento, critérios, indicadores, comunicam sobre a avaliação de forma mais fácil e mais precisa;
- Proporcionam um enquadramento que dá forma às relações entre clientes/comissários, diretores/gestores de programas e equipes de avaliação e esclarecem as funções de cada um;
- Servem como manuais de treinamento de avaliação, que dão clara orientação do caminho aos novatos e permitem que peritos em avaliação possam avaliar o seu grau de profissionalismo e identificar as necessidades de desenvolvimento;
- Funcionam como guias práticos sobre como fazer uma avaliação, passo a passo, conforme suas fases principais (definição de perguntas e propósitos da avaliação, recolhimento e valorização da informação empírica, comunicação e relato das descobertas).²²

Há também, claro, argumentos na literatura contra o uso dos parâmetros, mas pode-se rebatê-los. Ao discutir se os padrões são impeditivos para a inovação, por exemplo, deve-se salientar que, pelo contrário, os padrões devem promover o debate e a negociação, criando alternativas.

Algumas diretrizes podem soar, a princípio, contraditórias, como construir a viabilidade política e produzir relatórios imparciais. O avaliador depende da aprovação do contratante, mas não pode ser refém de suas vontades. Beywl (2004) afirma que aqueles “são critérios vagamente definidos, deixando espaço para interpretação, negociação e concretização, dependendo da finalidade específica da avaliação em questão e do contexto situacional.” Os dois parâmetros não serão alcançados em 100%, mas conforme a realidade permita. Associa-se a esse ponto outra crítica, a de que os parâmetros não são igualmente importantes para todas as audiências e experiências. Mas, assim como os relatórios devem se adaptar ao público ao qual se destinam, também os padrões o devem fazer.

Pode-se argumentar que não há qualquer obrigação dos governantes ou funcionários públicos para trabalhar com os resultados da avaliação padronizada, logo, sua utilidade é limitada. Mas

[a]pós vencer o desafio de tornar explícito o que se pretende alcançar, o processo de avaliação se reforça na medida em que superamos os desafios que restam: a eliminação da diferença qualitativa/quantitativa, a combinação de indicadores e informações necessárias para o processo de avaliação e a introdução de estratégias e incentivos para garantir os fluxos de informação relevante e a promoção da sua utilização. À medida que avançarmos nessas áreas poderá ser construído um processo de avaliação que estimula e apóia o

²² Tradução própria. Original em inglês.

processo de gestão e que alimenta o processo decisório com informação valiosa.²³ (MOKATE, 2002).

Deve ficar claro que as normas de avaliação esforçam-se para promover a utilização pelos tomadores de decisão, administradores públicos e políticos. A ampliação do uso depende, contudo, da mudança cultural nas organizações.

4.2 Padrões da JCSEE: uma visão geral

Os padrões ou *standards* dividem-se em quatro grupos temáticos: Utilidade, com sete quesitos (U1 – U7), Viabilidade, com três quesitos (F1 - F3), Adequação Ética, com oito quesitos (P1 – P8) e Rigor ou Exatidão, com doze quesitos (A1 - A12)²⁴.

A seguir os objetivos de cada grupo serão descritos resumidamente e, assim como feito nos trabalhos da JCSEE e da SEVAL, cada diretriz será numerada, receberá um título curto e será descrita em uma frase.

Cabe observar que os padrões individuais, bem como os quatro grandes grupos, são intencionalmente não ponderados, para que não se possa fazer nenhuma declaração preliminar sobre a importância particular de um padrão determinado ou grupo temático. Os pesos devem ser estabelecidos para cada avaliação, de acordo com a relativa importância dos quesitos, a atividade em foco (condução, definição, planejamento ou conclusão da avaliação, coleta ou análise das informações, construção dos relatórios ou do orçamento da avaliação) e conforme o julgamento dos avaliadores. Já para o estudo de caso que se segue, um exercício exploratório, não será atribuído peso aos quesitos.

4.2.1 Padrões de Utilidade

Segundo Rodrigues (2009):

Os *standards* de utilidade procuram assegurar que a avaliação é usada, o que supõe a resposta às necessidades de informação dos utilizadores, de modo claro e em tempo útil, bem como uma “orientação construtiva”, que conduza à recomendação, planificação (implementação, inclusivamente) acompanhamento e avaliação de “*ações de follow-up*”, destinadas a consolidar e desenvolver os aspectos fortes e/ou a eliminar, corrigir e melhorar os aspectos fracos, assim configurando o “impacto da avaliação”.

A utilidade da avaliação é conquistada pelo atendimento dos quesitos descritos no QUAD. 1. Essas diretrizes requerem que os avaliadores se familiarizem com as suas audiências, as

²³ Tradução própria. Original em espanhol.

²⁴ Os eixos serão tratados com seus nomes traduzidos do inglês para português, ainda que isso possa representar eventual perda de significado, por restringir demais a interpretação. No entanto, para manter a comparabilidade com o modelo original, o código alfanumérico associado a cada *standard* no documento do JCSEE foi mantido.

definem claramente, verifiquem as carências informacionais das mesmas, planejem a avaliação para responder a essas carências e relatem as informações relevantes em tempo hábil.

Quadro 1: Padrões de Utilidade

	Quesito	Definição
U.1	Identificação dos <i>stakeholders</i>	Pessoas envolvidas ou afetadas pela avaliação deveriam ser identificadas, para poderem ter suas necessidades respondidas.
U.2	Credibilidade do avaliador	O condutor da avaliação deveria ser confiável e reunir as competências para realizar a avaliação, de modo que os resultados dessa alcancem a máxima credibilidade e aceitação.
U.3	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 <i>stakeholders</i> envolvidos.
U.4	Identificação dos valores	As perspectivas, procedimentos e a racionalidade usada na interpretação das conclusões da avaliação deveriam ser cuidadosamente descritas, para estabelecer claramente as bases dos julgamentos de valor.
U.5	Clareza dos relatórios	Os relatórios da avaliação deveriam descrever detalhadamente o programa sendo avaliado, incluindo o contexto, os propósitos, procedimentos e conclusões da avaliação, de modo que a informação essencial seja provida e facilmente entendida.
U.6	Conclusão tempestiva dos relatórios e sua disseminação	Descobertas significativas no íterim da avaliação e relatórios de avaliação deveriam ser disseminados para os usuários interessados, para poderem usá-los em tempo oportuno.
U.7	Impacto da Avaliação	Avaliações deveriam ser planejadas, conduzidas e reportadas de modo a encorajar os <i>stakeholders</i> a acompanhá-las e utilizá-las, de modo que a preferência pelo uso da avaliação aumente.

Fonte: Elaboração própria.

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 daqueles que a tomarão para si.

4.2.2 Padrões de Viabilidade

Os padrões de exequibilidade (QUAD. 2) reconhecem que as avaliações definem e consomem recursos valiosos e que são geralmente executadas em um ambiente natural e não em laboratórios. Assim sendo, 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 compõe.

Os *standards* de exequibilidade pretendem garantir que a avaliação tem condições políticas e materiais para ser implementada como planejado e pretendido, o que exige diplomacia e procedimentos que não estorvem ou

interfiram com as actividades educativas (objecto de avaliação), que sejam práticos e que possam mobilizar os apoios e recursos necessários. (RODRIGUES, 2009).

Quadro 2: Padrões de Viabilidade

	Quesito	Definição
F.1	Procedimentos pragmáticos	Os procedimentos da avaliação deveriam ser práticos, para minimizar as interrupções enquanto a informação necessária é obtida.
F.2	Viabilidade política	A avaliação deveria ser planejada e conduzida com antecipação das posições dos diferentes grupos de interesse, para obter a cooperação destes, e tentativas de qualquer um desses grupos de sabotar as operações da avaliação, ou enviesá-la, ou desorientar os resultados poderiam ser evitadas ou contrabalanceadas.
F.3	Relações custo-benefício	A avaliação deveria ser eficiente e produzir informação de valor suficiente, de modo que os recursos despendidos possam ser justificados.

Fonte: Elaboração própria.

Os três padrões dessa categoria, tomados em conjunto, expressam se a avaliação é realista, prudente, diplomática e econômica.

4.2.3 Padrões de Adequação Ética

A avaliação pode afetar muitas pessoas, de forma muito variada, mas não pode violar o direito e o bem-estar delas. Para facilitar a proteção dos direitos dos indivíduos que se relacionam com a avaliação é que se estabeleceram padrões de adequação ética (QUAD. 3). Eles alertam contra a ilegalidade, a ação inescrupulosa e sem ética, ou contra ações despropositadas por aqueles que realizam as avaliações.

Quadro 3: Padrões de Adequação Ética

	Quesito	Definição
P.1	Orientação ao serviço	As avaliações deveriam ser desenhadas de modo a ajudar as organizações a endereçar e servir efetivamente as necessidades de todos os potenciais beneficiários do programa.
P.2	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.
P.3	Direitos da pessoa humana	Avaliações deveriam ser desenhadas e conduzidas de modo a respeitar e proteger os direitos e o bem-estar dos indivíduos.
P.4	Interação humana	Nas interações com outras pessoas, durante a avaliação, os avaliadores deveriam respeitar a dignidade humana e a auto-estima dos demais, de modo que seus participantes não sejam ameaçados ou sofram danos.

P.5	Avaliação Completa e Justa	A avaliação deveria ser completa e justa no exame e registro dos pontos fortes e fracos do programa sendo avaliado, de modo que as forças possam ser aumentadas e as áreas problemáticas abordadas.
P.6	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.
P.7	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.
P.8	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.

Fonte: Elaboração própria.

Essas diretrizes requerem que os avaliadores conheçam e sigam as normas referentes à privacidade, liberdade da informação e proteção dos seres humanos. Elas exigem que o processo de avaliação respeite os direitos de todos os envolvidos e de todos os afetados por seus resultados.

4.2.4 Padrões de Rigor ou Exatidão

Esses padrões determinam se uma avaliação produziu boas e relevantes informações. A avaliação de um programa deve ser abrangente, isto é, os avaliadores devem considerar o máximo de características de identificação do programa e devem recolher dados sobre as características julgadas especificamente importantes para a avaliação do mérito do programa. Além disso, as informações devem ser tecnicamente adequadas e os juízos devem estar explicitamente justificados pelos dados.

Os parâmetros de rigor ou exatidão (QUAD. 4) objetivam garantir que a avaliação irá revelar e transmitir informações precisas sobre o mérito do programa, que produzirá e disseminará informações utilizáveis e válidas.

Quadro 4: Padrões de Rigor ou Exatidão

	Quesito	Definição
A.1	Documentação do programa	O programa sob avaliação deveria ser descrito e documentado claramente e com precisão, de modo que o programa seja identificado claramente.
A.2	Análise do contexto	O contexto no qual o programa se insere deveria ser examinado com aprofundamento suficiente, de modo que suas possíveis influências no programa possam ser identificadas.
A.3	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.

A.4	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.
A.5	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.
A.6	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.
A.7	Informação sistematizada	A informação coletada, processada e reportada na avaliação deveria ser sistematicamente revisada e todos os erros encontrados deveriam ser corrigidos.
A.8	Análise da informação quantitativa	Em uma avaliação, a informação quantitativa deveria ser apropriadamente e sistematicamente analisada, de tal modo que as questões da avaliação sejam efetivamente respondidas.
A.9	Análise da informação qualitativa	Em uma avaliação, a informação qualitativa deveria ser apropriadamente e sistematicamente analisada, de tal modo que as questões da avaliação sejam efetivamente respondidas.
A.10	Conclusões justificadas	As conclusões assumidas em uma avaliação deveriam ser explicitamente justificadas, para que os diversos grupos de interesse possam avaliá-las.
A.11	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.
A.12	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: Elaboração própria.

Explicado o modelo desenvolvido pelo JCSEE, com seus quatro grupos temáticos e os parâmetros qualitativos que constituem cada um deles, torna-se possível realizar um exercício exploratório, qual seja, um estudo de caso para o estado de Minas Gerais.

5. Estudo de caso: a avaliação do programa Poupança Jovem

Algumas considerações antes de proceder ao estudo de caso. Em primeiro lugar, cabe lembrar que não é objetivo desse trabalho julgar se o Programa Poupança Jovem é pertinente ou relevante para a sociedade mineira. Um programa governamental, sob a ótica da avaliação, uma vez implantado, pode ser tomado como válido, isto é, presume-se que foi socialmente discutido e é socialmente aceito, cabendo à avaliação apenas medir em qual grau alcança os resultados que prometeu alcançar.

Em segundo lugar, o modelo do JCSEE não é tomado como o único ou o melhor compêndio de padrões a seguir no desenvolvimento de uma avaliação. Decerto todo avaliador considera uma série de requisitos antes de considerar concluída a sua avaliação. Parâmetros análogos aos do modelo da JCSEE podem estar presentes e outros, inclusive ignorados pelo modelo adotado nessa pesquisa, podem aparecer. A vantagem, no entanto, de se estabelecer explicitamente os *standards* já foi discutida, é uma forma de profissionalizar o trabalho, torná-lo replicável e passível de julgamentos fundamentados.

Em terceiro lugar, o estudo de caso é um exercício exploratório. Como toda avaliação, depende de critérios e valores subjetivos do avaliador. As avaliações nunca devem ser tomadas como leituras definitivas da realidade. Mas podem, oportunamente, influenciar o sistema ou o programa avaliado, como leituras parciais, sucessivas, melhorando a si mesmas e aos programas.

5.1 A importância da educação como foco de políticas públicas

Antes de explicar o Programa Poupança Jovem pretende-se discutir, brevemente, a motivação para políticas educacionais, especialmente no nível médio. A educação é muitas vezes tomada como panacéia, como solução para todos os problemas sociais. Todavia, é difícil acreditar que um único modelo, ou que uma única política pública, possa ser capaz de compreender toda a complexidade de problemas que alcança a sociedade brasileira e, a partir daí, corrigi-los.

Via de regra, assume-se que investimentos em educação são basilares na construção da nação que desejamos. Isto porque, embora a educação não seja, sozinha, a resposta para todos os conflitos e debates no nosso encaço, ela é, alternativamente, parte importante da resposta, na medida em que opera em favor da redução das desigualdades de oportunidade. A redução das desigualdades, segundo Barros, Mendonça e Henriques (2000), deve ser a prioridade estratégica, de modo a tornar possível erradicar a pobreza no país. Assim, a educação traz

consigo externalidades positivas, capazes de reduzir a pobreza, promover o crescimento econômico, a coesão social e a democracia.

A redução da pobreza pode ser um dos ganhos decorrentes do aumento dos anos de estudo e da qualidade do conhecimento adquirido nas escolas. O pobre brasileiro, mesmo quando trabalha, em função de não possuir ativos com valor financeiro, ganha pouco e permanece à margem do sistema creditício. Sua incapacidade de oferecer garantias, de endividar-se, torna-o incapaz de financiar o acesso aos ativos que poderiam libertá-lo de sua pobreza (KERSTENETZKY, 2002). O meio factível de prover acesso a estes ativos perpassa a atuação de políticas redistributivas, como a educação pública.

No Brasil o acesso ao ensino fundamental está praticamente universalizado, ao ensino médio, não. Em Minas Gerais, de acordo com a Pesquisa Nacional por Amostra de Domicílios – PNAD – levada a campo em 2008, 95,5% da população de sete a 14 anos freqüentava o ensino fundamental (IBGE, 2009).

Considerando a educação privada, no entanto, percebe-se que as crianças da classe média e alta geralmente entram nas escolas muitas vezes antes dos seis anos de idade²⁵. E permanecem por mais tempo. Crianças com pais menos abastados, além de terem menos contato com textos impressos do que crianças com pais ricos, tendem a adiar a entrada na escola até seus estudos tornarem-se passíveis de financiamento público. E podem deixar a escola mais cedo, como parece apontar a taxa de frequência líquida do jovem de 15 a 17 anos no ensino médio. Em Minas Gerais, em 2008, pela PNAD, esse indicador alcançava 56% (IBGE, 2009). Ou seja, 44% dos jovens no estado nessa faixa etária, aproximadamente 484.000 pessoas, estava fora do ensino médio. Dos que frequentavam, 86% o faziam na rede pública, evidenciando a importância da ação estatal no fornecimento desse serviço educacional. Outro fator que corrobora para a análise é o número médio de anos de estudo conforme a renda *per capita*, ou o estoque de escolarização, que tem grande variação em Minas Gerais. Para pessoas com mais de 25 anos de idade, em Minas Gerais, do quinto com menor renda *per capita* para aquele com maior, tem-se as seguintes médias: 4,6 - 5,2 - 5,4 - 6,8 - 9,9 anos de estudo (IBGE, 2009). O quinto mais rico tem, na média, mais que o dobro de anos de estudo que o quinto mais pobre.

[Para o Brasil as] desigualdades nesse acesso são pronunciadas, especialmente quando considerada a renda das famílias desses alunos [...]: no 1º quinto, 29,6% dos jovens de 15 a 17 anos cursam o ensino médio, proporção que para os mais ricos

²⁵ Essa afirmação pode ser feita a partir do texto do MEC, de 2004, citado nas referências, que atesta que as crianças da classe média e alta, aos seis anos, já estão inseridas no sistema de ensino regular.

(5º quinto) é de 78,5%, ou seja, a oportunidade de acesso ao ensino médio dos adolescentes mais ricos é mais de duas vezes a dos mais pobres. Os aspectos regionais também mostram diferenças. A situação é ainda pior entre os jovens pobres residentes da zona rural, onde apenas 23,8% estão incluídos no ensino médio. Jovens de cor negra, sejam do 1º quinto ou do 5º quinto, estão em desvantagem em relação aos brancos. (IPEA, 2009a).

Pelo estoque de anos de estudo, por quintos de renda, percebe-se que apenas o quintil mais rico alcançou a escolarização mínima obrigatória, de oito anos de estudos, estabelecida pela Constituição Federal de 1988. Políticas educacionais focalizadas, com o objetivo de aumentar o acesso e a permanência dos alunos nas escolas, inclusive no ensino médio, quando o limite inferior de oito anos de estudo já foi superado, podem contribuir para diminuir essa desigualdade no cenário educacional do país.

[No Brasil,] além disso, para os jovens dessa faixa etária [15 a 17 anos] que estão cursando o ensino médio repetem-se os problemas de fluxo escolar, o que acarreta taxas médias esperadas de conclusão também inadequadas [...]. A estimativa é de que apenas 66,6% dos alunos que ingressam no ensino médio o concluem. (IPEA, 2009a).

É preciso aumentar a taxa de conclusão do ensino médio, bem como ampliar o acesso ao ensino superior e a conclusão dessa etapa. Na faixa etária seguinte, entre 18 e 24 anos, apenas 13,9% dos brasileiros, segundo a PNAD 2009 (IBGE, 2009a) cursava a educação superior, ou seja, parcela mínima da população com idade ideal para esse nível de ensino consegue chegar efetivamente à educação superior. Esse fato se deve “aos entraves observados no fluxo escolar do ensino fundamental e médio, que tem elevada taxa de evasão e baixa taxa média esperada de conclusão” (IPEA, 2009a).

Sob este ponto de vista, a ampliação do alcance do ensino público é questão de direito, é uma conquista social que amplia o leque de oportunidades dos nascidos mais pobres e favorece a probabilidade de romperem o ciclo de pobreza. Segundo Marshall (1967), o direito do cidadão adulto de ter sido educado nos remete ao direito da criança de frequentar a escola. Como direito, torna-se obrigação do Estado para com os cidadãos em formação. Destarte, a política de educação básica e a avaliação de sua eficácia respondem a um imperativo de justiça.

A oportunidade de adquirir as competências cognitivas essenciais à vida contemporânea, a exemplo do letramento e do desenvolvimento do raciocínio lógico-matemático, constitui um **direito**. Desse ângulo, a educação compõe uma das faces do direito do indivíduo à autonomia e à igualdade de oportunidades, porque não há indivíduo autônomo sem os recursos cognitivos necessários à expressão de suas idéias, sentimentos e interesses, nem se pode falar de igualdade de oportunidades em um cenário de grandes assimetrias cognitivas. Em uma

palavra, a educação diz respeito ao direito à **igual liberdade**²⁶. (MINAS GERAIS, 2002).

A importância do *background* familiar já foi discutida por Mare (1979) e estudada no caso brasileiro por Pastore e Valle Silva (1999), que concluíram que o nível de educação e o status da profissão do pai, bem como o local de moradia (rural ou urbana), são preditores da idade de ingresso no mercado de trabalho e da quantidade de anos de estudo acumulada pelos filhos. Filhos de médicos, por exemplo, profissão de status alto, começariam a trabalhar, em média, com 21 anos de idade, enquanto filhos de trabalhadores rurais analfabetos, profissão de status baixo, começariam a trabalhar com 9,7 anos.

Essa importância foi revisitada, mais recentemente, por Rios-Neto, Riani e César (2002 e Riani e Rios-Neto (2008). Essas análises abordam as motivações para a escolha do nível de investimento em educação, mostrando, principalmente, em que termos se dá a estratificação educacional. Pais pobres e com nível educacional mais baixo geralmente transferem baixa escolaridade para os filhos, reforçando o ciclo de pobreza e a baixa escolaridade. Segundo Barros *et al.* (2001) “a escolaridade dos pais, e em particular a da mãe, é, de forma robusta, a mais importante para determinar o desempenho educacional dos jovens em questão”. O determinismo na tendência de reprodução das desigualdades sociais na formação de capital humano de sucessivas gerações pode, contudo, ser contrariado pela atuação da escola (FERRÃO; FERNANDES, 2003).

Outra externalidade positiva da educação é a coesão social, derivada, dentre outros fatores, das expectativas que os indivíduos formam. Existem expectativas de posição, que

[...] trata-se de desigualdades que se repetem historicamente, entre grupos sociais, étnicos, de certas localidades, de gênero, de tal modo que ter nascido em um determinado grupo, local, etnia ou gênero revela-se o melhor preditor das chances de “sucesso” de um indivíduo dentro da sociedade, ou da quantidade de opções reais diante dele. (KERSTENETZKY, 2002).

Em contraste, existem as expectativas legítimas, que podem ser instigadas pela educação, embora a promoção da equidade não seja simples. Essas expectativas dependem da igualdade moral estabelecida, ou igualdade de oportunidades, e podem flexibilizar o sucesso alcançado por cada indivíduo, de forma meritocrática, recompensando os esforços e punindo as negligências dos cidadãos. Se o ambiente institucional é composto por sujeitos que reconhecem, em si mesmos, idênticos valores individuais dentro da sociedade, idênticos pesos

²⁶ Grifos no original.

no trato com a Administração Pública, a tendência é de se acatar as normas e regras impostas, de se promover a incorruptibilidade, de se respeitar a democracia.

Por tudo isso, mostra-se importante investir recursos públicos em zonas de vulnerabilidade, principalmente no campo educacional, com o intuito de diminuir a desigualdade e facilitar que os indivíduos rompam o ciclo intergeracional da pobreza.

5.2 O Programa Poupança Jovem

Antes de proceder ao estudo de caso em si é pertinente apresentar o Programa Poupança Jovem, sua justificativa, objetivo geral e específico, o histórico e o público-alvo do programa.

O Programa Poupança Jovem (PJ) é um programa de cunho social, com foco integral em alunos do ensino médio, da rede estadual, residentes em municípios ou regiões consideradas vulneráveis, ou seja, que possuem elevadas taxas de evasão escolar, além de altos índices de violência ou baixo Índice de Desenvolvimento Humano (IDH). O programa visa promover a capacitação dos jovens, pela permanência e conclusão do ensino médio, pela realização de atividades profissionalizantes e que aumentem seu protagonismo na sociedade. Além das aulas normais são lecionadas aulas de reforço e acompanhamento. Os estudantes têm também acesso a programas extracurriculares, como cursos de idiomas, esportes e formação de liderança comunitária.

A implantação do programa buscou justificar-se na atuação do Estado como agente promotor da ruptura do ciclo intergeracional de pobreza. Uma família pobre tende a continuar pobre, pois são muitos os empecilhos para superar essa condição, alguns deles associados à baixa quantidade e qualidade da escolaridade obtida. O programa, que fornece capital financeiro, através de uma bolsa concedida aos beneficiários graduados no ensino médio, e capital humano, através do ensino secundário e profissionalizante, visa permitir aos jovens as condições para se emanciparem da trajetória intergeracional.

O objetivo geral do PJ advém do Decreto Estadual nº. 44.548, de 22 de junho de 2007, que contém o regulamento do programa e afirma que ele “tem por finalidade o desenvolvimento de ações que busquem oferecer aos beneficiários a oportunidade de desenvolvimento pessoal e social, aumentar a taxa de conclusão do ensino médio e reduzir os índices de criminalidade entre os jovens”. Seus objetivos específicos são aumentar a taxa de conclusão do ensino médio; propiciar conhecimentos e habilidades gerais e específicas para a inserção no mercado de trabalho; e desenvolver competências e habilidades gerais, bem como valores como o protagonismo, a responsabilidade social e o cooperativismo.

O Poupança Jovem tem como eixos a educação geral, a educação para o trabalho e a cultura. Tem-se, assim, coerência com a Lei Nacional de Diretrizes e Bases da Educação, que entende o jovem sob três perspectivas: da pessoa, do cidadão e do futuro profissional. As políticas públicas, sob este ponto de vista, devem orientar seus beneficiários para a formação pessoal, a preparação para o exercício da cidadania plena e para a qualificação profissional.

Deseja-se diminuir a participação no mercado de trabalho dos jovens entre 15 e 17 anos. Defende-se que

[...] nesta fase da vida é fundamental postergar a entrada no mercado de trabalho para viabilizar, sobretudo, a permanência na escola e a conclusão do ensino médio com qualidade. Este período corresponde ao que vem sendo chamado de moratória social: um crédito de tempo que permite ao jovem protelar as exigências sociais típicas da vida adulta, especialmente relativas ao casamento e ao trabalho, e lhe possibilita um maior contato com experiências e experimentações que podem favorecer o seu pleno desenvolvimento, não apenas em termos de formação educacional e aquisição de treinamento e capacitação, mas também em termos de outras vivências típicas que fazem parte da sociabilidade juvenil. (IPEA, 2009b)

A meta do Programa Poupança Jovem é ter, até o final de 2010, 50 mil jovens beneficiários (até 2009 foram beneficiados em torno de 32 mil jovens²⁷, em oito cidades mineiras: Ribeirão das Neves, Sabará, Esmeraldas, Ibirité, Governador Valadares, Montes Claros, Teófilo Otoni e Juiz de Fora).

A contrapartida exigida dos alunos é a frequência mínima a 75% das aulas e aprovação em todas as disciplinas, bem como nenhum envolvimento com a criminalidade, de acordo com a regulamentação prevista no Decreto.

Nos dois primeiros anos o jovem, se aprovado e em conformidade com os requisitos impostos pelo termo de compromisso, pode retirar cem reais, a título de adiantamento, dos mil reais a ele atribuídos pela conclusão de cada ano escolar. No total, para os três anos do ensino médio, são depositados em uma conta corrente três mil reais para cada beneficiário, dois mil e oitocentos se sacados os cem reais do primeiro e segundo ano, corrigidos monetariamente. Ao formar, o estudante recebe o volume financeiro acumulado e é livre para escolher como utilizá-lo.

Ademais, o PJ não atua desarticulado de outras políticas estaduais. Pelo contrário, está associado a programas educacionais, como a distribuição de livros didáticos gratuitos,

²⁷ Informações retiradas do site do programa: <http://www.poupancajovem.mg.gov.br>.

programas de segurança pública, como o programa Fica Vivo, e de desenvolvimento regional, como o Projeto Travessia²⁸.

5.3 Aplicação do modelo da JCSEE à avaliação do Programa Poupança Jovem

A avaliação do Programa Poupança Jovem será avaliada segundo o modelo do JCSEE, a partir de dados secundários, como o contrato entre o governo de Minas Gerais e o CEDEPLAR, de prestação de serviços técnicos envolvendo a avaliação de impacto de projetos estruturados na área de educação, entre eles o PJ²⁹. Além disso, serão avaliados os cinco relatórios parciais entregues pelo CEDEPLAR, responsável pela avaliação de impacto que nos interessa. O primeiro trata do plano amostral e do questionário da avaliação do Programa Poupança Jovem, o segundo e o terceiro tratam da operação de campo, o quarto da metodologia para a avaliação de impacto e o quinto faz a primeira avaliação de impacto do Programa Poupança Jovem em uma perspectiva longitudinal.

Cumprе salientar que a avaliação do Poupança Jovem foi tratada nos contatos dos avaliadores com os alunos e com as diretorias das escolas como uma pesquisa de avaliação e acompanhamento dos alunos do ensino médio, intitulada “PESQUISA JOVEM: avaliação do ensino médio”, nunca como uma pesquisa vinculada ao Programa PJ. As razões para adoção dessa prática foram, principalmente, não criar constrangimento aos alunos dos municípios que aderiram ao programa e nem a sua consequente vinculação a uma avaliação pessoal enquanto, ao mesmo tempo, evitavam-se especulações juntos aos alunos dos municípios que não eram alvo do programa Poupança Jovem, mas que fariam parte do grupo de controle.

Como a avaliação contratada é de impacto, ela não necessariamente deveria abordar todos os pontos que uma avaliação global alcança. A especificidade do impacto, no entanto, já discutida, justifica porque esse tipo de avaliação foi escolhida, em detrimento das demais, de processos e de custo-benefício, por exemplo. Pretende-se, com a avaliação de impacto, testar os elos causais que levaram à implantação do programa político, verificando os resultados que tem gerado para os seus beneficiários e o impacto social do projeto, ou, como dito anteriormente, mensurar os efeitos líquidos da intervenção.

Cada grupo temático do modelo de *standards* será avaliado separadamente, buscando nos documentos pesquisados subsídios que justifiquem o pleno atendimento aos quesitos que o

²⁸ Esses e outros projetos e programas do governo de Minas Gerais podem ser encontrados no site: <http://www.geraes.mg.gov.br>.

²⁹ O referido contrato é o 631/08, celebrado entre a SEPLAG e a Fundação de Desenvolvimento da Pesquisa da Universidade Federal de Minas Gerais.

compõem. Quando não houver evidências do atendimento ao item, serão descritas as informações que foram procuradas e não foram encontradas, evidenciando porque naquele ponto a avaliação foi considerada incompleta.

5.3.1 Padrões de Utilidade

O primeiro parâmetro de utilidade é a identificação dos *stakeholders*, que são todos os atores que tenham algum tipo de interesse na política, programa ou projeto. Widmer, Landert e Bachmann (2000) listam alguns grupos que deveriam constar como partes do processo avaliativo: aqueles que tomarão decisões sobre o futuro do objeto sendo avaliado; aqueles responsáveis pela concepção e estruturação do objeto avaliado; os envolvidos com a implementação do programa; os grupos de tratamento e os grupos envolvidos no ambiente no qual se inserem, entre outros.

Pode-se afirmar que a avaliação do PJ buscou apoio e contribuição de diversos grupos. Os membros do governo, clientes da avaliação, foram os primeiros a apontar que informações consideravam essenciais para a avaliação, no escopo da contratação do projeto. A equipe responsável pela execução do programa foi consultada antes da elaboração do questionário e com ela foram discutidos os pontos principais que deveriam constar no instrumento de investigação (CEDEPLAR, 2007). O trabalho de campo consultou os alunos beneficiários e futuros participantes do grupo de controle, para diferenciar seus atributos e contextos (CEDEPLAR, 2008a). Isso após o devido contato com a Secretaria de Estado de Educação e Secretaria de Desenvolvimento Social, que coordenam a pesquisa em conjunto com o CEDEPLAR (CEDEPLAR, 2007).

Os membros do referido Centro, responsáveis pela interpretação dos dados coletados, e a equipe de campo, responsável pela coleta, atendem ao requisito de credibilidade do avaliador. Para serem tomados como confiáveis pelas várias partes afetadas, as seguintes características são particularmente importantes para os avaliadores: ter integridade pessoal, mostrar independência e demonstrar competência comunicativa e social (WIDMER, LANDERT e BACHMANN, 2000). O contrato firmado entre o governo e o CEDEPLAR estabeleceu requisitos para a composição da chefia das equipes, pesquisadores com doutorado ou pós-graduação em andamento nas áreas afins da pesquisa. Os demais membros das equipes, como os aplicadores dos questionários, tinham no mínimo graduação em andamento (CEDEPLAR, 2008a, 2008b). As equipes, portanto, detinham formação acadêmica e, além disso, foram treinadas nos instrumentos de coleta e registro, para conquistar a massiva participação dos

alunos avaliados e evitar atritos na coleta ou erros no preenchimento e construção da base de dados (CEDEPLAR, 2008a).

No planejamento da avaliação é necessário estabelecer qual informação é indispensável para responder às questões colocadas, seu alcance e seletividade. Devem-se diferenciar as informações meramente interessantes ou desejáveis, no fim, desnecessárias, daquelas essenciais e indispensáveis. Na “PESQUISA JOVEM: Avaliação do Ensino Médio” uma multiplicidade de pesquisas anteriores foi consultada como forma de orientação para cada tema desenvolvido (CEDEPLAR, 2007). Como exemplos, foram examinados, em relação à saúde sexual e reprodutiva, questionários como os da DHS (*Demography and Health Survey*), assim como alguns *surveys* internacionais, como o “*Los Angeles Family and Neighborhood Survey*” para questões referentes ao uso de drogas, saúde e violência.

O modelo do JCSEE chama atenção para as fontes disponíveis para coleta de dados. Na avaliação em análise, o Censo Escolar, realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais do Ministério da Educação (INEP/MEC), forneceu um banco de dados com informações detalhadas sobre as escolas da rede estadual nos municípios de interesse (CEDEPLAR, 2007). Já o questionário aplicado aos alunos matriculados no ensino médio da mesma rede foi o alicerce para a construção da base de dados sobre o grupo tratamento e sobre o grupo de controle (CEDEPLAR, 2007).

Por outro lado, o ato de avaliar exige que sejam feitos juízos de valor. O avaliador utiliza modelos teóricos e orientações valorativas no seu processo de interpretação. Logo, “se o processo interpretativo pretende ser convincente, inteligível e avaliável, é imperativo tornar transparente quais foram as bases que motivaram os julgamentos de valor”³⁰ (WIDMER, LANDERT e BACHMANN, 2000). Os relatórios do CEDEPLAR apontam com clareza como foi selecionada a amostra de tratamento e de controle e os procedimentos estatísticos usados. As amostras foram selecionadas “tendo com base o procedimento de amostragem por conglomerados em um estágio com estratificação geográfica” (CEDEPLAR, 2007). As metodologias, quantitativas principalmente, como o pareamento pelo escore de propensão, utilizado para estimar os efeitos do Programa Poupança Jovem (CEDEPLAR, 2009a), são apresentadas brevemente, com sugestões de consulta bibliográfica para maior detalhamento.

O modelo de *standards* recomenda que a linguagem dos relatórios seja precisa e adequada ao público destinatário. Os relatórios avaliados são todos destinados ao órgão contratante da

³⁰ Tradução própria. Original em inglês.

avaliação, o governo mineiro e, neles, os termos importantes são explicitamente definidos e usados com consistência, com destaque para os procedimentos estatísticos e as definições derivadas da econometria. Informações importantes são sumarizadas em tabelas e gráficos, formas recomendadas de comunicar as considerações e descobertas, especialmente as estatísticas descritivas do grupo controle (CEDEPLAR, 2008b) e grupo tratamento (CEDEPLAR, 2008a), bem como os resultados da avaliação de impacto (CEDEPLAR, 2009b). Além do relatório na forma escrita, os resultados foram expostos em encontros entre a equipe executora da avaliação e os membros do Estado para Resultados, que apóiam a coordenação exercida pelos servidores da SEPLAG. Entretanto, seminários e palestras, com o intuito de disseminar os resultados e as metodologias utilizadas, poderiam ter sido realizados e registrados nos relatórios e não o foram. No entanto, cumpre observar a cláusula de confidencialidade do contrato, que tornam as informações e os dados objetos de sigilo, até 1º de fevereiro de 2011, quando passarão a ser considerados domínio público.

O sigilo dessas informações é justificado por questões legais e contratuais, bem como fundamentado pela necessária articulação política e técnica, no sentido de que os objetivos e metas de um não são independentes do outro. Sob o ponto de vista técnico, no entanto, a disseminação das informações geradas pela avaliação deveria alcançar o maior número de interessados possível, ampliando a capacidade contributiva da avaliação para a tomada de decisão e para a retroalimentação do programa. Sob o ponto de vista político, os ganhos do sigilo parecem ser estimados de forma a superar os ganhos da transparência, comprometendo os ganhos potenciais da divulgação dos resultados da avaliação. Há um *trade-off*, um conflito de escolhas: a resolução de uma parte do problema, a manutenção do sigilo, acarreta em outro problema, a menor divulgação dos resultados.

Quesito importante, a avaliação perde muito da sua capacidade contributiva se não for tempestiva, isto é, se seus prazos não permitirem o uso das descobertas para melhoria do programa ou para municiar a tomada de decisão. Relatórios parciais são formas de realizar esses comunicados ainda durante a avaliação, particularmente se puderem reorientar o programa para a obtenção dos resultados desejados. A avaliação de quatro relatórios parciais e de um quinto, com a primeira avaliação longitudinal, apontavam que esse quesito tinha sido atendido. No entanto, a entrevista com o servidor alocado no Programa Estado para Resultados apontou que os prazos estabelecidos para as entregas foram extrapolados e que, por isso, houve maior dificuldade de usar seus resultados para retroalimentar o programa.

Destarte, a importante condição prévia, que ajuda a promover ou realizar a expectativa de impacto da avaliação, de envolver os diversos atores relevantes no planejamento e organização do projeto de avaliação foi atendida de forma satisfatória, dadas as evidências já comentadas. Também foi considerada útil a comunicação de relatórios de progresso claros e frequentes ao contratante. A comunicação ao público geral, embora desejada, encontra um obstáculo significativo e justificado na confidencialidade dos dados. No entanto, a entrevista com o servidor do EPR apontou que essa divulgação é desejada e também que todas as solicitações feitas pelo CEDEPLAR para acesso ao banco de dados da avaliação do Programa Poupança Jovem foram autorizadas, assim como a realização de pesquisas acadêmicas com os mesmos dados. O objetivo da avaliação é traçar o perfil dos jovens que aderiram ao programa e, também, avaliar o impacto do Poupança Jovem na vida desses jovens durante sua trajetória escolar. Um objetivo técnico, não político. Tornar os resultados sigilosos enquanto se realiza um pleito para o governo mineiro, na transição entre 2010 e 2011, foi uma forma cautelosa de tratar da delicada interseção entre o ciclo político e a continuidade das atividades tecnocratas.

5.3.2 Padrões de Viabilidade

Esse grupo temático recomenda que os procedimentos sejam pragmáticos, com mínimas interrupções no andamento do programa para coleta das informações. “Os procedimentos da avaliação são as ações particulares tomadas no processo de coletar e usar a informação para julgar o valor e o mérito do programa”³¹ (JCSEE, 1994). Entre eles incluem-se, mas não somente, como são feitos os contratos, como as fontes de dados são escolhidas, como os dados e a informação são recolhidos, gravados, arquivados, recuperados, como os dados são analisados e como as descobertas são reportadas. Os procedimentos escolhidos devem ser aqueles que minimizam as interrupções e que se adéquam ao tempo, orçamento e equipe disponíveis. Portanto, é importante deixar claro quais as vantagens e desvantagens dos métodos escolhidos, assim como podem e não podem contribuir.

Na avaliação do PJ o número de escolas amostradas foi escolhido levando-se em conta a eficiência amostral e o orçamento disponível, conforme relatado no primeiro relatório, que descreve o plano amostral e o questionário (CEDEPLAR, 2007). Além disso, as visitas às escolas, para aplicação dos questionários, foram agendadas previamente, conforme a disponibilidade das turmas, e o preenchimento do questionário pelos alunos durava em torno de uma hora, durante o período escolar, mas em um intervalo curto, sem comprometer uma

³¹Tradução própria. Original em inglês.

parcela considerável de tempo dos alunos e professores (CEDEPLAR 2008a, 2008b). Ações de destaque estão associadas ao emprego de digitadores experientes para transcrição dos questionários e ao uso de software livre para elaboração da máscara de entrada dos dados, o CSPro (*Census and Survey Processing System*), cuja elaboração e verificação é acelerada pelas disponibilidades da aplicação computacional (CEDEPLAR, 2008a, 2008b).

Já para minimizar resistências à avaliação pode ser evidenciada a preocupação, concernente ao contrato, de não publicizar informações e resultados antes do próximo período eleitoral. Além disso, potenciais reações que enviesariam os dados, como a dos gestores das escolas ou dos alunos, se avisados que a pesquisa tinha como foco o Programa Poupança Jovem, foram evitadas, pela relação da pesquisa ao acompanhamento genérico de jovens no ensino médio (CEDEPLAR, 2007). A construção da viabilidade política consistiu, também, da incorporação das equipes das Secretarias Estaduais de Educação e de Desenvolvimento Social na gestão da pesquisa, atores relevantes tanto para a consecução dos dados quanto para a implementação de mudanças a favor do aprimoramento do programa (CEDEPLAR, 2007).

Uma avaliação deve também ser efetiva para ser considerada viável. Para isso seus benefícios devem igualar ou exceder os seus custos. Custos entendidos no seu valor total, isto é, custo social e monetário da avaliação, de todas as fontes físicas e humanas necessárias à prática avaliativa. Os benefícios também são os sociais totais derivados da avaliação, como otimização dos efeitos do programa, possíveis economias de custos e aumento da aceitação do programa. No entanto, quantificar esses efeitos totais para a sociedade não é trivial, muito menos para uma avaliação ainda em andamento. Ainda que complexas, essas estimativas devem compor os relatórios da avaliação e, no caso em estudo, não foram encontradas menções ao assunto. O foco são os resultados da avaliação, executada com autonomia, mas os processos, embora não prioritários em relação aos resultados, não devem ser deixados de fora.

5.3.3 Padrões de Adequação Ética

Em relação à responsabilidade fiscal, importante padrão de adequação ética, as ações a serem executadas durante a avaliação deveriam ter seus custos considerados e o orçamento alocado de forma correspondente. Esse cuidado com o orçamento se propõe a evitar o desperdício de recursos e/ou a realização incompleta do trabalho contratado. Como se trata de uso de recursos públicos, a contabilidade deveria ser meticulosa, com prestação de contas por parte dos contratados conforme estipulado no contrato. No entanto, essa prestação não estava presente nos contratos ou nos relatórios parciais avaliados.

As avaliações deveriam monitorar a efetividade do programa, ou seja, se ele entrega o valor público prometido, sendo, pois, orientada ao serviço. Deveriam também incluir a perspectiva dos vários *stakeholders* na avaliação, para atendê-los conforme suas expectativas e necessidades, bem como manter a avaliação atualizada em relação ao estado da arte do programa implantado. Como evidenciado, a avaliação do PJ cumpriu esses requisitos.

Por essa ótica, contudo, revela-se importante disponibilizar a informação para o público geral, respondendo a uma concepção mais ampla de *accountability*, para favorecer a democracia, ou o poder de decisão e de governança do povo. Cumprirá, portanto, após o período de sigilo contratual, dar ampla divulgação à pesquisa e seus resultados, fortalecendo o programa, modificando-o ou descartando-o. Haverá sempre razões para tornar públicas as informações coletadas na avaliação, mas não a qualquer custo. Como ressalta o modelo, “os clientes da avaliação devem ser encorajados a prover informação apropriada a todas as partes afetadas pela avaliação, tempestivamente, em uma forma linguística adequada a cada audiência”³² (JCSEE, 1994).

Para ser adequado eticamente o relacionamento do avaliador com aqueles que encomendaram a avaliação deve ser pautado pelo mútuo respeito e confiança. Esse é o ambiente adequado para a elaboração de um contrato escrito formal, que contenha as mais importantes tarefas e regras que cabem a cada uma das partes. O objetivo da contratualização é diminuir a chance de desentendimentos e facilitar a solução desses quando ocorrerem. Os contratos devem prever as regras em relação, no mínimo, a: “financiamento, prazos, pessoas envolvidas, relatórios a serem produzidos ou publicados, conteúdo, metodologia e procedimentos a serem seguidos” (JCSEE, 1994). Com o passar do tempo torna-se evidente que revisões são necessárias, para as quais podem existir renegociações das cláusulas estabelecidas. Conforme comentado anteriormente, houve contratualização formal da avaliação e o documento reuniu as regras mínimas: a forma de pagamento da despesa, o prazo de execução (prorrogado por um ano por aditamento), a formação do corpo técnico, a frequência dos relatórios.

A avaliação também respeitou os direitos dos envolvidos, os baseados na lei e os baseados na ética, senso comum e cortesia, como recomendado pelo modelo do JCSEE. Os avaliadores devem conhecer e respeitar, pelo bem da moral e por razões pragmáticas, todos os direitos daqueles que são alcançados pela avaliação. A eventual violação desse quesito poderia levar a sanções legais e talvez criar oposição ou descrédito para a avaliação. Não foi o que se observou para a avaliação do PJ. Por exemplo, o contato com as escolas foi formal, por

³² Tradução própria. Original em inglês.

telefone e carta, agendando a visita e solicitando a participação dos alunos na pesquisa (CEDEPLAR, 2008a, 2008b). Por sua vez, o anonimato dos alunos que responderam aos questionários, cadastrados no sistema por códigos, foi assegurado e os alunos foram avisados dessa garantia antes de participar (CEDEPLAR, 2007). Destarte, os avaliadores se “precaveram contra os efeitos potencialmente ameaçadores ou nocivos de suas interações com os participantes”³³ (JCSEE, 1994). O modelo aponta que as pessoas que sentem sua dignidade ou auto-respeito violados não apenas perdem criatividade potencial, mas podem vir a agir de uma maneira que prejudique a avaliação. Boas relações interpessoais e habilidades de comunicação foram vitais para o cumprimento adequado da avaliação do Programa Poupança Jovem.

Ademais, a avaliação tem como um dos seus objetivos reconhecer as forças do programa avaliado, bem como oportunidades de melhoria. Por outro lado, pode buscar salientar as fraquezas e as eventuais ameaças à obtenção dos resultados esperados. O quinto relatório avaliado, a análise de impacto propriamente dita, apontou algumas das relações de causa-efeito testadas e as respostas a elas (CEDEPLAR, 2009b). No entanto, dada a confidencialidade do relatório até fevereiro de 2011, não serão tecidos comentários sobre as descobertas, exceto que apontam, como desejado pelo modelo, forças e fraquezas, consequências pretendidas e não pretendidas.

Já conflitos de interesse podem surgir em várias situações. Os avaliadores podem, mas não devem, em determinadas situações, ficar sujeitos a qualificar um programa de forma enviesada, dada a preocupação em receber futuros contratos com o mesmo cliente. Não foi observado esse viés na avaliação analisada. Os processos, a metodologia, os resultados, todos foram passíveis de discussão prévia e detalhada também nos relatórios.

5.3.4 Padrões de Rigor ou Exatidão

A documentação do programa é o primeiro padrão de apuro. O programa não será plenamente avaliado a não ser que seja completamente discutido na avaliação. Recomenda-se explicitar exatamente o que é estudado, pois, se aprofundado, o estudo do programa pode permitir que se façam conexões entre o objeto da avaliação e os efeitos explorados, ou ajudar a identificar efeitos perversos não previstos. Particularmente, pode-se perceber as diferenças entre a concepção e a implementação do programa, levando à aprendizagem em relação ao funcionamento do mesmo, como ocorreu com a avaliação do Programa Poupança Jovem, que

³³ Tradução própria. Original em inglês.

contém uma explicação do escopo, do objetivo, dos marcos referenciais, da legislação reguladora e do público-alvo do programa (CEDEPLAR, 2007).

O contexto do programa - que inclui “sua localização geográfica, o tempo no qual toma forma, o ambiente político e social que o cerca, as atividades que competem com ele, a equipe que o coordena, as condições econômicas”³⁴ (JCSEE, 1994) - foi avaliado. Fatores contextuais muitas vezes influenciam decisivamente os efeitos atingíveis por uma avaliação e, se bem fundamentados, podem tornar possível estimar para quais condições as conclusões podem ser generalizadas. As inferências estatísticas realizadas nos relatórios, a partir da amostra selecionada, puderam ser realizadas apenas para a população de referência descrita no plano amostral (CEDEPLAR, 2008a, 2008b). A situação sócio-econômica dos alunos que compuseram o grupo controle e de tratamento também foi considerada (CEDEPLAR, 2008a, 2008b).

Os objetivos a perseguir, as questões a responder e os procedimentos escolhidos foram documentados ao longo da avaliação, dando transparência ao processo avaliativo. A documentação dos procedimentos incluiu, entre outros, a descrição detalhada da organização, da coleta e do processamento dos dados, bem como da sua análise e comunicação (CEDEPLAR 2007, 2008a, 2008b).

Um dos padrões de rigor estabelece que as fontes de informação utilizadas na avaliação devem ser defensáveis. Trata-se de explicar a origem e os limites das fontes de informação, de modo que o público possa, por si mesmo, decidir se a informação é de boa qualidade ou não. A credibilidade da avaliação poderia ser criticada se não houvesse descrição das fontes de dados ou se elas fossem incompletas. Mas não o são no caso em estudo e as conclusões da avaliação podem ser julgadas pertinentes ou não, justificadas ou não, dada a profundidade e a abrangência dos dados disponíveis e utilizados, cuja coleta e tratamento foram descritos nos relatórios (CEDEPLAR, 2007, 2008a, 2008b).

A validade é uma diretriz que consiste na “razoabilidade das inferências decorrentes dos resultados encontrados no processo de obtenção de informação”³⁵ (JCSEE, 1994). A validade responde se e como uma medida reflete exatamente o conceito que ela pretende medir. É a adequação da informação à metodologia. Já a confiabilidade, por sua vez, trata da consistência ou estabilidade da qualidade da medida, seja entre diferentes instrumentos de medição, ou por ter diferentes pessoas envolvidas, ou pelo decorrer do tempo. A validade e a

³⁴ Tradução própria. Original em inglês.

³⁵ Tradução própria. Original em inglês.

confiabilidade de uma medida estão intimamente relacionadas, e na seleção ou avaliação de formas alternativas de coletar os dados, ambas as qualidades devem ser tratadas.

Na avaliação do Programa Poupança Jovem houve constante preocupação com a validade e confiabilidade, sendo recorrentes procedimentos de checagem e verificação. Os questionários preenchidos pelos alunos foram conferidos e respostas conflitantes ou rasuradas foram confirmadas junto aos respondentes (CEDEPLAR, 2008a, 2008b). A lista de respondentes foi cruzada com a lista de chamada da escola, para diferenciar homônimos e evitar conflitos no registro dos indivíduos (CEDEPLAR, 2008a, 2008b). O programa computacional utilizado para transcrição dos questionários utilizava máscara de preenchimento, para evitar erros, e realizava testes automáticos de consistência, por exemplo, entre a idade e data de nascimento do respondente (CEDEPLAR, 2008a, 2008b). Já para o tratamento dos resultados, foram apresentados cálculos com diferentes metodologias quantitativas, com o objetivo de prover robustez à análise e às descobertas (CEDEPLAR, 2009b).

Essas verificações e checagens tiveram como objetivo minimizar os erros e garantir a qualidade e consistência dos dados. Eram possíveis erros simples, como digitação errada de entradas em sistemas, e erros complexos, como interpretações equivocadas dos dados disponíveis. Métodos apropriados foram usados para revisar a informação, como testes de consistência e validação com os responsáveis pelo fornecimento da informação (CEDEPLAR, 2008a, 2008b), mecanismos de controle descritos nos relatórios.

As conclusões da avaliação são baseadas na análise qualitativa e quantitativa. Ambas devem seguir o rigor metodológico do modelo empregado, explicando suas limitações e a significância dos resultados. A avaliação estudada teve sua metodologia detalhadamente descrita (CEDEPLAR, 2009a), bem como suas consequências criticamente examinadas (CEDEPLAR, 2009b). Novamente, os relatórios se mostraram suficientemente claros e elucidativos, de modo que aquele que os receber será capaz de decidir se suas conclusões são corretas e adequadas.

Por fim, as avaliações podem receber muitas críticas, se não forem bem justificadas. Podem falhar e levar a decisões ruins. Podem extrapolar, negativamente, as expectativas a ela relacionadas. Para evitar essas dificuldades a avaliação pode ser estudada, conforme recomenda o último parâmetro de rigor ou exatidão, ou seja, pode ser desenvolvida uma meta-avaliação, levada a cabo pela própria equipe ou por um consultor externo. Como restrições orçamentárias são comuns, nem sempre recompensa fazer uma profunda meta-avaliação. Mas auto-avaliações dos que participaram do processo de construção da avaliação do programa

não deveriam ser omitidas, posto que encerram oportunidades de aprendizado, conhecimento que pode ser transferido para avaliações futuras. Assim, o relativamente baixo volume de capital necessário para realizar essa atividade deveria constar desde o orçamento inicial, mas não parece ter sido o caso, visto que não há comentários a esse respeito nos relatórios avaliados. A entrevista com o servidor do EPR demonstrou que há espaço para a realização formal dessas auto-avaliações por parte da equipe que ficou responsável pela “PESQUISA JOVEM: avaliação do ensino médio”. O estudo de caso aqui desenvolvido, contudo, ainda que exploratório e restrito, buscou contribuir dessa forma.

6. Considerações finais

O desenvolvimento do Programa Estado para Resultados, de forma incremental, demonstra que a gestão por resultados, inclusive em Minas Gerais, não tem caráter fixo e inalterável. Pelo contrário, deseja-se que seja variável, que seja capaz de se adaptar ao contexto, de aperfeiçoar-se e ao ambiente que o governa. Decerto o programa não pode, na contramão, enrijecer-se, nem adquirir hábitos associados a instituições burocráticas improdutivas, como buscar a maximização do seu orçamento como um fim em si mesmo.

O modelo de gestão por resultados em Minas Gerais é recente ainda, visto que foi implantado em 2007, portanto com muitas características em desenvolvimento e carentes de maturação. Sistemas de monitoramento e avaliação precisam ser instaurados incrementalmente. Leva tempo para reunir perícia, para criar sistemas de dados robustos, para conduzir as avaliações, portanto, o sucesso imediato no uso das avaliações não é usual.

O programa Estado para Resultados tem caráter provisório, encerrando-se, tal como foi concebido, em janeiro de 2011. No entanto, quer seja renovado ou terminantemente extinto, o conhecimento e a capacidade produtiva desenvolvidos no âmbito do programa não deveriam ser perdidos. Para evitar que sua herança seja descartada também naquela data, sem se considerar como e em que medida o que fora realizado contribuiu para o desenvolvimento da gestão pública mineira, está em andamento, dentro das atividades do programa, a agenda de melhorias. Trata-se de um levantamento, com consulta a órgãos e atores considerados importantes no cenário estadual, entre secretarias do governo, representantes de carreiras estratégicas, empresários e sociedade civil, dos avanços dos últimos anos e com propostas de alternativas para o futuro próximo.

A adoção de alguns instrumentos e orientações associadas à gestão pública para resultados contribuiu para a profissionalização do serviço público no estado mineiro e apontou potenciais espaços para melhorias. Espaços para atuação estatal em prol da eficácia e eficiência das políticas públicas e programas governamentais, bem como em favor da *accountability*, entendida em suas várias formas. Entre esses instrumentos destaca-se o uso e a disseminação das avaliações.

Ao executar esse trabalho percebeu-se que a avaliação já se mostra importante em Minas Gerais, mas ainda não se converteu em um processo indispensável na gestão pública, assim como apontado por Mokate (2002) em relação aos países ao redor do mundo. Como apontado, a autora afirma que parte desse problema tem como causa a cultura gerencial enraizada, que

vê na avaliação um “monstro” e não um associado. Mudanças na cultura gerencial, por conseguinte, podem facilitar e promover o uso das avaliações como instrumentos aliados à gerência. Essas mudanças têm, para ela, quatro características principais, que podem balizar o amadurecimento do modelo mineiro. Primeiro, a cultura deve ser modificada de modo a permitir o equívoco e o ajuste. A estrutura organizacional deve permitir a flexibilidade e a inovação, passíveis de erros. Então, as reorientações nas políticas, em busca das transformações e resultados desejados, tornariam a avaliação indispensável, pois as descobertas desta apontariam o caminho da inovação dinâmica. Em segundo lugar, o contexto gerencial também deveria promover a autoridade daqueles que recebem as informações, no sentido de poderem utilizá-las dentro das organizações e requererem as modificações apontadas como necessárias. É preciso permitir, em função dos objetivos existentes, a atuação sobre a informação. Em terceiro lugar, o contexto deveria promover incentivos organizacionais e gerenciais favoráveis à utilização da informação, sejam implícitos ou explícitos, em forma de remuneração variável ou reconhecimento e prestígio, por exemplo. Finalmente, mas não menos importante, o ambiente ótimo para o desenvolvimento das avaliações deveria possuir mecanismos de ajustes. Isso implica formas que “permitam a realocação de recursos financeiros, físicos e humanos; implica a autoridade para redefinir estratégias operacionais e/ou modificar as características do bem, serviço ou informação que se está oferecendo, para alcançar os objetivos desejados”³⁶ (MOKATE, 2002).

E como dito anteriormente, essas mudanças gerenciais não precisam, necessariamente, preceder o uso das avaliações. Essas podem ser utilizadas como ferramentas para a promoção daquelas mudanças gerenciais. Em Minas Gerais, por exemplo, algumas constatações apontam essa possibilidade como o caminho escolhido.

Primeiro, a ampliação da participação de instituições do governo estadual, como a Fundação João Pinheiro, no rol de organizações aptas a promover a avaliação de políticas públicas, reforça as vantagens de internalizar as técnicas e a capacidade de avaliar do próprio governo, além de favorecer o sistema pelo aumento da comunicação entre os órgãos e suas equipes:

[Q]uatro projetos (Viva Vida, Pro-Hosp, Minas Sem Fome e Integração da Gestão da Segurança Pública) foram selecionados para avaliação da Fundação João Pinheiro (FJP), dado seu interesse em importar essa tecnologia, optando por fazer um processo de *learning by doing*, por meio da contratação de especialistas em avaliações para tutoriar equipes de avaliação da própria instituição. (GUIMARÃES; CAMPOS; 2009)

³⁶ Tradução própria. Original em espanhol.

Essa citação foi extraída de um artigo no qual Eder Campos e Tadeu Guimarães, assessor e coordenador executivo do Programa Estado para Resultados, respectivamente, apontam que Minas Gerais precisa avançar na ligação entre o orçamento e o recente sistema de monitoramento e avaliação. Segundo eles “ainda há muito a se desenvolver para traçar o elo entre as iniciativas realizadas, e por consequência os recursos alocados, e os resultados e impactos gerados” (GUIMARÃES; CAMPOS, 2009). Entre essas iniciativas, citam as avaliações de impacto em curso, que podem justificar a tomada de decisões e a continuidade ou interrupção de cada programa avaliado.

Cumprido ressaltar que dificilmente uma avaliação será perfeita e maximizará todas as dimensões de desempenho, ou atenderá a todos os quesitos considerados ideais. Persegui-los, no entanto, é fundamental, no intuito de tornar a realidade cada vez mais próxima da utopia. Assim, para diminuir a variabilidade das avaliações, para ampliar o uso das avaliações e para aumentar a capacidade das secretarias e do próprio governo de supervisionar os avaliadores externos - alguns dos muitos benefícios apontados pelos defensores dos padrões de avaliação - uma alternativa seguida por diversas Sociedades de Avaliação, na Europa e na América do Norte, por exemplo, mas não somente, foi a adoção dos *standards*.

Foi com o modelo originado nos Estados Unidos da América, do *Joint Committee on Standards for Education Evaluations*, que se inaugurou uma metodologia posteriormente adaptada por diversas instituições correlatas. Esse modelo, apresentado ao longo do trabalho, reúne características consideradas ideais na composição de uma avaliação, às vezes chamadas de diretrizes, outras vezes de estandartes ou normas. Mas vale lembrar que elas não são absolutas, porque buscam considerar as nuances dos contextos, as especificidades dos governos e das realidades nas quais esses governos atuam.

Esses parâmetros foram reunidos em quatro grupos temáticos: Utilidade, Viabilidade, Adequação Ética, Rigor ou Exatidão. Enfatize-se que a avaliação é, por natureza, uma atividade complexa e sistêmica. Mas a separação dos padrões em grupos, de forma funcional, organiza os subsídios que orientam a avaliação da qualidade. Da qualidade, cuja promoção e defesa é o objetivo do modelo, seja das avaliações dos serviços educacionais, quando o modelo é aplicado nesse campo, seja nas demais áreas de atuação governamental.

O estudo de caso elaborado no presente trabalho foi acerca de uma avaliação de serviço educacional, sobre a avaliação do Programa Poupança Jovem, intitulada “PESQUISA JOVEM: avaliação do ensino médio”. O QUAD. 5 resume as considerações acerca da mesma. No quadro, os parâmetros considerados plenamente atendidos são realçados em verde.

Padrões cujo atendimento foi parcial foram realçados em amarelo e em vermelho aparecem os padrões considerados não atendidos.

Quadro 5: Resumo da avaliação da “PESQUISA JOVEM: avaliação do ensino médio”

Utilidade	Viabilidade	Adequação Ética	Rigor ou Exatidão
1. Identificação dos <i>stakeholders</i>	1. Procedimentos pragmáticos	1. Orientação ao serviço	1. Documentação do programa
2. Credibilidade do avaliador	2. Viabilidade política	2. Acordo formal	2. Análise do contexto
3. Alcance e Seleção da Informação	3. Relações custo-benefício	3. Direitos da pessoa humana	3. Propósitos e procedimentos descritos
4. Identificação dos valores		4. Interação humana	4. Fontes de informação defensáveis
5. Clareza dos relatórios		5. Avaliação Completa e Justa	5. Informação válida
6. Conclusão tempestiva dos relatórios e sua disseminação		6. Divulgação de resultados	6. Informação confiável
7. Impacto da Avaliação		7. Conflito de interesses	7. Informação sistematizada
		8. Responsabilidade Fiscal	8. Análise da informação quantitativa
			9. Análise da informação qualitativa
			10. Conclusões justificadas
			11. Relatório Imparcial
			12. Meta-avaliação

Fonte: Elaboração própria.

No eixo Utilidade considerou-se que a pesquisa atende quase todos os parâmetros que o compõem, como credibilidade do avaliador contratado e identificação e consideração das preferências dos *stakeholders*. Como ponto de melhoria, a manutenção dos prazos acordados se revela ponto importante para permitir que os resultados dos relatórios sejam apropriados em tempo hábil pelo ciclo de formulação e reformulação da política. A avaliação identificou os impactos do programa e ela, por si mesma, poderá gerar impacto quando forem expostos seus resultados, quando superado o prazo de confidencialidade que restringe essa divulgação atualmente. Seja pela divulgação de relatórios para o público em geral, seminários ou palestras. Cabe destacar uma consideração associada a esse ponto: um alto grau de transparência sobre como as avaliações tratam os objetos avaliados e sobre suas descobertas poderá ampliar a credibilidade, a integridade e a propriedade da prática avaliativa como uma profissão respeitada.

No grupo temático Viabilidade a avaliação do Programa Poupança Jovem foi considerada apropriada, porque apresentou procedimentos pragmáticos e mostrou-se atenta à viabilidade política. Em relação às relações custo-benefício, de difícil mensuração, não foram encontradas considerações nos documentos analisados. Estimar os benefícios gerados pela avaliação é um problema muito discutido na literatura, não consensual, mas que também não deve ser ignorado. No entanto, posto que a avaliação do PJ ainda está em andamento, espera-se que futuros relatórios parciais ou o relatório final venham preencher essa lacuna.

Sobre Adequação Ética, assim como nos dois grupos temáticos anteriores, a avaliação do Programa Poupança Jovem respondeu positivamente a grande maioria dos quesitos do modelo do JCSEE. A orientação ao serviço, entretanto, reforça a necessidade de dar ampla divulgação aos resultados, alcançando, principalmente, os beneficiários do programa e os demais grupos interessados no mesmo. Em relação ao quesito responsabilidade fiscal, considerada a autonomia do prestador de serviço, a contabilidade do dispêndio na avaliação poderá vir a compor os documentos que descrevam o seu histórico, de modo a se confirmar o uso adequado dos recursos públicos.

O último grupo temático, Rigor ou Exatidão, teve todos os parâmetros considerados respeitados. Destarte, a avaliação do Programa Poupança Jovem atende positivamente os quatro grandes eixos e pode ser tomada como *benchmark* para as avaliações futuras e em desenvolvimento concomitante. Uma consideração importante, no entanto, é feita em relação ao quesito meta-avaliação. Aos profissionais envolvidos na elaboração e consecução da avaliação do PJ é recomendável a realização de uma auto-avaliação, que permita a documentação das lições aprendidas e a generalização dessas práticas para outros trabalhos e pesquisas.

Referências

- ABRUCIO, Fernando. Trajetória recente da gestão pública brasileira: um balanço crítico e a renovação da agenda de reformas. **RAP Revista de Administração Pública**, Edição Especial Comemorativa 1967-2007, 2007.
- ALA-HARJA, Marjukka; HELGASON, Sigurdur. Em direção às melhores práticas de avaliação. **Revista do Serviço Público**, Brasília, v. 51, n. 4, out./dez. 2000. Brasília, ENAP, 2000.
- BAKER, Judy L. **Evaluating the impact of development projects on poverty: a handbook for practitioners**. Washington, D.C., 2000.
- BARROS, Ricardo Paes de; MENDONÇA, Rosane; HENRIQUES, Ricardo. Desigualdade e pobreza no Brasil: retrato de uma estabilidade inaceitável. **Revista Brasileira de Ciências Sociais**, v. 15, n. 42, 2000.
- BARROS, Ricardo Paes de; MENDONÇA, Rosane; SANTOS, Daniel Domingues dos; QUINTAES, Giovani. Determinantes do desempenho educacional no Brasil. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 31, n. 1, 2001.
- BEHELAINÉ, Cinthia Helena de Oliveira. **A gestão pública por resultados no estado de Minas Gerais: um estudo de caso sobre o Programa Estado para Resultados**. Monografia (graduação) – Fundação João Pinheiro, Escola de Governo de Minas Gerais. Belo Horizonte, 2009.
- BERMAN, Paul. *Designing implementation to match Policy Situation: A contingency analysis of programmed and adaptative implementation; The Rand Corporation*, P6211. 1978.
- BEYWL, Wolfgang. *The Role of Evaluation in Democracy: Can it be strengthened by Evaluation Standards? A European Perspective*. In: **Journal of MultiDisciplinary Evaluation**, 6, 2004.
- BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil e Códigos**. Barueri, SP: Manole, 2003.
- BRASIL. Ministério da Educação. **Ensino Fundamental de nove anos: orientações gerais**. 2004.
- BRESSER-PEREIRA, Luiz Carlos. “Burocracia pública e Reforma Gerencial”. **Revista do Serviço Público** de 1937 a 2007, número comemorativo dos 70 anos. Brasília: ENAP, 2007.
- CAVALCANTI, Mônica Maria de Arruda. **Avaliação de Políticas Públicas e Programas Governamentais - uma abordagem conceitual**. Interfaces de Saberes (João Pessoa), v. 6, 2006. Disponível em: <<http://interfacesdesaberes.fafica.com/seer/ojs/include/getdoc.php?id=34&article=24&mode=pdf>>. Acesso em 07 de mai. de 2010.
- CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL DE MINAS GERAIS – CEDEPLAR. **Relatório: plano amostral e questionário**. Pesquisa Poupança Jovem. Produto 1. Belo Horizonte, 2007.
- CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL DE MINAS GERAIS – CEDEPLAR. **Relatório: operação de campo**. Pesquisa Poupança Jovem. Produto 2. Belo Horizonte, 2008a.

CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL DE MINAS GERAIS – CEDEPLAR. **Relatório: final campo 2008**. Pesquisa Poupança Jovem. Produto 3. Belo Horizonte, 2008b.

CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL DE MINAS GERAIS – CEDEPLAR. **Relatório 4: Metodologia para a Avaliação de Impacto dos Programas de Educação em Minas Gerais**. Pesquisa Poupança Jovem. Belo Horizonte, 2009a.

CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL DE MINAS GERAIS – CEDEPLAR. **Relatório Ribeirão das Neves: avaliação de impacto, 2007-2008**. Pesquisa Poupança Jovem. Produto 5. Belo Horizonte, 2009b.

CENEVIVA, Ricardo. **DEMOCRACIA, ACCOUNTABILITY E AVALIAÇÃO**: a avaliação de políticas públicas como mecanismo de controle democrático. 2005. 161 f. Dissertação (Mestrado) - Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo, 2005.

CORRÊA, Izabela Moreira. Planejamento estratégico e gestão pública por resultados no processo de reforma administrativa do Estado de Minas Gerais. **Revista de Administração Pública**. Rio de Janeiro, v. 41, n. 3, maio-junho 2007.

COSTA, Valeriano Mendes Ferreira. A dinâmica institucional da reforma do Estado: um balanço do período FHC. In: ABRUCIO, Fernando Luiz e LOUREIRO, Maria Rita Garcia (orgs.). **O Estado numa era de reformas: os anos FHC – Parte 2**. Brasília: MP/SEGES, 2002.

ENAP. **Glossário** - Acompanhamento, Monitoramento e Avaliação dos Programas e Projetos Sociais do Fundo Nacional de Desenvolvimento da Educação (FNDE). 2007.

FERRÃO, M. A.; FERNANDES, C.O. O efeito-escola e a mudança – Dá para mudar? Evidências da Investigação Brasileira. **Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 1, n. 1, 2003.

GUIMARÃES, Tadeu Barreto; CAMPOS, Eder Sá. Gestão por Resultados em Minas Gerais: Análise Crítica do Arranjo Incremental de Monitoramento e Avaliação. XXXIII EnANPAD 2009. XXXIII Encontro da ANPAD. São Paulo: ANPAD, 2009.

IAIES, G. Evaluar las evaluaciones. In: **Evaluar las Evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa**. IPE – UNESCO, Buenos Aires, Argentina, 2003.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Síntese de indicadores sociais. Uma análise das condições de vida da população brasileira. 2009. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/indicador_minimos/sinteseindicadores2009/indic_sociais2009.pdf>. Acesso em: 20 de dez. de 2009.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA – IPEA. **Comunicado da Presidência nº 32** - PNAD 2008: Primeiras Análises - Educação, Gênero e Migração. 2009a. Disponível em: <http://www.ipea.gov.br/003/00301009.jsp?ttCD_CHAVE=12435>. Acesso em: 18 de mai. de 2010.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA – IPEA. **Comunicado da Presidência nº 36** – PNAD 2008: Primeiras Análises - Juventude e Desigualdade racial. 2009b. Disponível em: <http://www.ipea.gov.br/003/00301009.jsp?ttCD_CHAVE=12933>. Acesso em: 18 de mai. de 2010.

JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1994). **The program evaluation standards**, 2ed edition. How to assess evaluations of educational programs. Thousand Oaks: Sage Publications. 1994.

KERSTENETZKY, Celia L. Por que se importar com a desigualdade. **DADOS - Revista de Ciências Sociais**, Rio de Janeiro, v. 45, n. 4, 2002.

MARE, Robert D. Social background composition and educational growth. **Demography**, v.16, n.1, Feb. 1979.

MARSHALL, T. H.. **Cidadania, classe social e status**. Rio de Janeiro: Zahar, 1967.

MINAS GERAIS. Secretaria de Estado de Educação. **Revista do professor**. A avaliação, o direito à educação e a desigualdade de oportunidades. 2002. Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd.

MOKATE, Karen Marie. *Eficácia, eficiencia, equidad y sostenibilidad: qué queremos decir?* Banco Interamericano de Desarrollo, Instituto Interamericano para el Desarrollo Social (Indes), 2001.

MOKATE, Karen Marie. *Convirtiendo el "monstruo" en aliado: la evaluación como herramienta de la gerencia social*. **Revista do Serviço Público**, Brasília, v. 53, n. 1, jan./mar. 2002.

NADER, Rosa Maria. **A avaliação como ferramenta para uma gestão pública orientada para resultados**. O caso do Governo Federal Brasileiro. Documento elaborado para o CLAD. 2005. Disponível em <<http://www.bresserpereira.org.br/view.asp?cod=2094>>. Acesso em: 02 de dez. de 2009.

O'DONNELL, Guillermo. Polyarchies and the (un)rule of law in Latin America: a partial conclusion. In: MENDES, Juan; O'DONNELL, Guillermo and PINHEIRO, Paulo Sergio(eds.). **The (un)rule of law and the underprivileged in Latin America**. Notre Dame: University of Notre Dame Press, 1999.

PASTORE, José; VALLE SILVA, Nelson do. **Mobilidade social no Brasil**. São Paulo: Macron Books, 2000.

PRZEWORSKI, Adam. "Sobre o Desenho do Estado: uma Perspectiva Agent X Principal". In: PEREIRA, Luiz Carlos; SPINK, Peter. **Reforma do Estado e Administração Pública Gerencial**. Rio de Janeiro: Fundação Getúlio Vargas Editora, 1998.

QUIRÓS, Mario. Responsabilización por el control de resultados. In: **La responsabilización en la nueva gestión pública Latinoamericana**. Buenos Aires: CLAD, 2000.

RIANI, J. L. R; RIOS-NETO, E. L. G. Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros?. **Revista Brasileira de Estudos de População**, vol.25, n.2, 2008.

RIOS-NETO, E. L. G. ; CÉSAR, Cibele Comini; RIANI, J. L. R. Estratificação educacional e progressão escolar por série no Brasil. **Pesquisa e Planejamento Econômico**. Rio de Janeiro, v. 32, 2002.

RODRIGUES, Pedro. Recensão das obras do *Joint Committee on Standards for Educational Evaluation* (1994, 2003, 2009). The program evaluation standards; The student evaluation standards & The personnel evaluation standards. Sísifo. **Revista de Ciências da Educação**, 9, 2009. Disponível em <<http://sisifo.fpce.ul.pt>>. Acesso em 20 de abr. de 2010.

SANO, Hironobu ; ABRUCIO, F. L. . Promessas e Resultados da Nova Gestão Pública no Brasil: o caso das Organizações Sociais de Saúde em São Paulo. RAE. **Revista de Administração de Empresas**, v. 48, 2008.

SERRA, Alberto. **Modelo Aberto de Gestão para Resultado no Setor Público**. Tradução de Ernesto Montes-Bradley y Estayes. Natal, RN: SEARH/RN, 2008.

SULBRANDT, José. A avaliação de programas sociais: uma perspectiva crítica dos modelos usuais. In: Kliksberg, Bernardo (org.). **Pobreza: uma questão inadiável**. Brasília: ENAP, 1994.

VILHENA, Renata *et al.* (Org.). **O Choque de Gestão em Minas Gerais**: políticas da gestão pública para o desenvolvimento. Belo Horizonte: Editora UFMG, 2006.

WEISS, Carol H. How can theory-based evaluation make greater headway? **Evaluation Review**, 1997.

WIDMER, Thomas; LANDERT, Charles; BACHMANN, Nicole. **Evaluation Standards of SEVAL, the Swiss Evaluation Society (SEVAL Standards)**. SEVAL, Berne. 2000. Disponível em <<http://www.seval.ch>>. Acesso em 20 de dez. de 2009.

APÊNDICE A – Questionário aplicado na entrevista

- 1) Os resultados da avaliação foram divulgados para quais grupos de pessoas?
- 2) Como foi a apropriação dos resultados da avaliação, até aqui, por parte do governo e da equipe executora do programa?
- 3) Como se deu essa divulgação? Comitês, encontros?
- 4) Superado o prazo estabelecido como limite temporal da confidencialidade da base de dados e da avaliação em si, há expectativas de que o CEDEPLAR ou os órgãos do governo darão ampla divulgação às descobertas da pesquisa?
- 5) A produção do CEDEPLAR envolve relatórios para os beneficiários e/ou para as escolas que ofertam o programa?
- 6) É possível dizer que houve conflitos de interesse durante a negociação da avaliação? Como eles foram articulados, em reuniões, telefonemas, encontros?
- 7) Há expectativa de que sejam estimados os benefícios gerados pela avaliação? Já é possível perceber se houve ganhos tais como otimização dos efeitos do programa, possíveis economias de custos ou aumento da aceitação do programa?
- 8) Houve prestação de contas por parte dos avaliadores? Há alguma discriminação da forma e/ou natureza dos dispêndios, como quantidade de horas trabalhadas por cada consultor e o custo dessas, consumo de material, etc.?
- 9) Haverá ou houve auto-avaliação por parte dos avaliadores do Programa Poupança Jovem? Algum tipo de relato de experiência ou lições aprendidas?