

Universidade de Brasília - UnB
Faculdade de Economia, Administração, Contabilidade e Ciência da
Informação e Documentação - FACE
Programa de Pós-Graduação em Administração
Curso de Mestrado Acadêmico

Selma Maria Hayakawa Cunha Serpa

Para que Avaliar?
Identificando a Tipologia, os Propósitos e a Utilização das Avaliações de
Programas Governamentais no Brasil

BRASÍLIA
2010

SELMA MARIA HAYAKAWA CUNHA SERPA

Para que Avaliar?
Identificando a Tipologia, os Propósitos e a Utilização das Avaliações de Programas
Governamentais no Brasil

Dissertação apresentada ao Programa de Pós -
Graduação em Administração – PPGA – da
Faculdade de Economia, Administração,
Contabilidade e Ciência da Informação e
Documentação (FACE) da Universidade de
Brasília (UnB) como requisito parcial à obtenção
do título de Mestre em Administração.

Orientador: Prof. Dr. Paulo Carlos Du Pin
Calmon

BRASÍLIA
2010

Serpa, Selma Maria Hayakawa Cunha.

Para que Avaliar? Identificando a Tipologia, os Propósitos e a Utilização das Avaliações de Programas Governamentais no Brasil/Selma Maria Hayakawa Cunha Serpa. – Brasília, 2010.

Orientador: Paulo Carlos Du Pin Calmon
Dissertação (Mestrado) – Universidade de Brasília, Programa de Pós-Graduação em Administração, PPGA.

SELMA MARIA HAYAKAWA CUNHA SERPA

Para que Avaliar?

Identificando a Tipologia, os Propósitos e a Utilização das Avaliações de Programas Governamentais no Brasil

Esta dissertação de mestrado foi julgada e aprovada como requisito parcial para obtenção do grau de mestre em Administração no Programa de Pós-Graduação em Administração da Universidade de Brasília.

Brasília, 05 de março de 2010.

Prof. PhD. Paulo Carlos Du Pin Calmon
Universidade de Brasília, PPGA.
Presidente

Prof^ª Dr^a Janann Joslin Medeiros
Universidade de Brasília, PPGA.
Membro Titular
Examinador Interno

Prof. Dr. Jorge Abrahão de Castro
Instituto de Pesquisa Econômica Aplicada, IPEA.
Membro Titular
Examinador Externo

Prof. Dr. Luiz Guilherme de Oliveira
Universidade de Brasília, PPGA
Suplente

RESUMO

Esta pesquisa trata do estudo da avaliação de programas governamentais sob a perspectiva da utilização dos seus resultados. A utilização tem sido um campo muito explorado internacionalmente, cujas pesquisas iniciaram no final dos anos 70, quando os avaliadores passaram a questionar a efetiva utilidade das avaliações para influenciar decisões e contribuir para o aperfeiçoamento de programas e políticas. No Brasil, os estudos ainda são incipientes. Com vistas a oferecer uma contribuição para a compreensão deste tema, principalmente quanto aos seus aspectos conceituais e práticos, este trabalho apresenta uma análise empírica dessas pesquisas aplicadas ao caso brasileiro, além de uma consolidação da literatura relativa às pesquisas realizadas, com destaque para aquelas que fundamentam a trajetória evolutiva do conhecimento atual sobre a utilização da avaliação.

Os principais trabalhos destacados na literatura se propuseram a analisar os fatores que influenciam a utilização dos resultados gerados pela avaliação. Para examinar a observância desses fatores, foram aplicados os padrões de avaliação de programas definidos pelo *Joint Committee on Standards for Educational Evaluation* a um conjunto de avaliações relativas aos programas do PPA 2004-2007, bem como a avaliações realizadas pelo TCU por meio das auditorias operacionais, com vistas a analisar em que medida as sistemáticas e abordagens metodológicas aplicadas favorecem a utilização dos resultados por elas produzidos. As avaliações de programas foram selecionadas a partir do estudo da institucionalização da função avaliativa no Brasil, por meio do qual foram identificados os tipos, os objetivos, as sistemáticas de implementação e estratégias metodológicas empregadas nas avaliações realizadas pelos órgãos governamentais em nível federal.

Esta pesquisa não teve a pretensão de generalizar os achados observados por meio de inferências estatísticas. Trata-se de um estudo baseado na pesquisa bibliográfica e no estudo de casos múltiplos, cuja coleta de dados foi realizada por intermédio do levantamento documental, entrevistas e aplicação de questionário, sendo que as respostas ao questionário foram apresentadas segundo uma escala de notas, cujas médias foram consideradas para evidenciar a observância dos padrões de avaliação definidos pelo *Joint Committee*.

Os resultados demonstraram que as avaliações de programas, no âmbito da administração pública federal, foram instituídas, principalmente, com vistas a subsidiar tanto o processo orçamentário e a melhoria da gestão, quanto o controle dos resultados das ações governamentais, atendendo, em ambos os casos pesquisados, ao propósito da *accountability*. A sistemática aplicada às avaliações de programas do PPA 2004-2007 pode ser tipificada como monitoramento. No que diz respeito às avaliações realizadas pelo TCU, caracterizam-se como auditorias operacionais. Com relação à abordagem metodológica, pode-se evidenciar que as avaliações dos programas do PPA 2004-2007 não foram concebidas e implementadas de forma a favorecer a utilização dos seus resultados pelos diversos atores interessados para atender a todos os objetivos propostos. Quanto às avaliações do TCU, por outro lado, a abordagem metodológica mostrou-se aderente aos padrões de avaliação, o que favorece a sua utilização.

Palavras-chave: Avaliação de programas governamentais. Utilização. Padrões de avaliação de programas. Avaliações dos programas do PPA. Avaliações de programas do TCU.

ABSTRACT

This research is a study on evaluation of government programs under the perspective of the utilization of its outcomes. Evaluation use is a field that has been extensively explored internationally, whose researches began by the end of the 70s, when evaluators started to question the effective utilization of evaluations to influence decisions and contribute for the improvement of programs and policies. In Brazil, the studies are still incipient. Because of this the objective of this study is to contribute to the comprehension on this issue, especially with regards to its conceptual and practical aspects. This work presents more than a consolidation of research literature in relation to the researches conducted, highlighting those that outline the evolution of modern knowledge about the utilization of evaluation, using empirical analysis of these researches applied to the Brazilian case. The main researches highlighted in the research literature proposed to analyze the factors that influenced the utilization of the results generated by the evaluation. To examine the observance of these factors, standards of program evaluation defined by the Joint Committee on Standards for Educational Evaluation were applied to a group of Multiannual Plan programs evaluations (PPA) and another group of evaluations conducted by the Brazilian Court of Audit (TCU) through performance auditing, to analyze to what level the systematics and methodological approaches favor the utilization of the results generated by them. The evaluations of programs were selected from the study of institutionalization of the evaluative function in Brazil, through which were identified the types, objectives, systematics of implementation and methodological strategies used in evaluations conducted by federal offices of government.

This research did not have the intention of generalizing the results observed through statistical inferences. This is a study based on the literature research and the study of multiple cases, of which the collection of data was performed by documental research, interviews and the use of questionnaires, being that the responses to the questionnaires were presented by a grade-scale, where the average grades were considered to make evident the observance of evaluation standards defined by the Joint Committee.

The results demonstrated that the evaluation of governmental programs, in the area of federal public administration, were instituted mainly to subsidize both the budgetary process and the betterment of administration, as well as to the control of the results of governmental actions, concerning both cases researched, with the purpose of accountability. The systematic applied to the evaluations of programs of the Brazilian Federal Multiannual Plan – PPA 2004-2007 - can be typified as monitoring. In what is said with respect to the evaluations performed by TCU, they are characterized as performance auditing. With regards to the methodological approach, it can be shown that evaluations from the Multiannual Plan 2004-2007 programs were not thought-out and implemented in ways to ease the use of their results by all stakeholders interested to address all the proposed objectives. The TCU evaluations, on the other hand, adhered its methodological approach to the standards of evaluation, which aids in its utilization.

Keywords: Government program evaluation. Utilization. Program evaluation standards. Multiannual Plan program evaluation. Brazilian Court of Audit program evaluation.

LISTA DE FIGURAS

FIGURA 1 - ESTRUTURA DA UTILIZAÇÃO DA AVALIAÇÃO DE COUSINS E LEITHWOOD (1986, P 348, tradução nossa)	58
FIGURA 2 - MODELO ANALÍTICO DE JOHNSON (1998).....	70
FIGURA 3 - MODELO TEÓRICO DE UTILIZAÇÃO DA AVALIAÇÃO DE JOHNSON (1998, P. 104, Tradução Nossa)	71
FIGURA 4 - MODELO LÓGICO DA AVALIAÇÃO	73
FIGURA 5 - MÉTODO DE ESTUDO DE CASO.....	95
FIGURA 6 - ATUAÇÃO DOS SISTEMAS DE CONTROLE INTERNO, EXTERNO E DO SISTEMA DE PLANEJAMENTO E ORÇAMENTO	112

LISTA DE GRÁFICOS

GRÁFICO 1 - MÉDIAS DOS PADRÕES DE UTILIDADE POR AVALIAÇÃO	153
GRÁFICO 2 - MÉDIAS DOS PADRÕES DE VIABILIDADE POR AVALIAÇÃO.....	155
GRÁFICO 3 - MÉDIAS DOS PADRÕES DE PROPRIEDADE POR AVALIAÇÃO	160
GRÁFICO 4 - MÉDIAS DOS PADRÕES DE ACURACIDADE POR AVALIAÇÃO.....	163

LISTA DE TABELAS

TABELA 1 - COMPARATIVO DO QUANTITATIVO DE PROGRAMAS VIGENTES PPA 2004-2007 E AVALIADOS NO PERÍODO.....	119
TABELA 2 - MÉDIAS DOS PADRÕES DE UTILIDADE POR AVALIAÇÃO	152
TABELA 3 - MÉDIAS DOS PADRÕES DE VIABILIDADE POR AVALIAÇÃO.....	154
TABELA 4 - MÉDIAS DOS PADRÕES DE PROPRIEDADE POR AVALIAÇÃO	156
TABELA 5 - MÉDIAS DOS PADRÕES DE ACURACIDADE POR AVALIAÇÃO	161

LISTA DE QUADROS

QUADRO 1 - DEMONSTRATIVO DA RELAÇÃO ENTRE OS PROPÓSITOS DAS AVALIAÇÕES SUGERIDOS POR WEISS (1998), PATTON (2008) E MARK ET AL. (2000).....	33
QUADRO 2 - PADRÕES DE UTILIDADE.....	36
QUADRO 3 - PADRÕES DE VIABILIDADE.....	36
QUADRO 4 - PADRÕES DE PROPRIEDADE.....	37
QUADRO 5 - PADRÕES DE ACURACIDADE.....	39
QUADRO 6 - DEFINIÇÕES DOS TERMOS UTILIZAÇÃO E INFLUÊNCIA SEGUNDO LAWRENZ <i>ET AL.</i> , 2009, TRADUÇÃO NOSSA.....	41
QUADRO 7 - TIPOLOGIA DE UTILIZAÇÃO SEGUNDO LAWRENZ <i>et al.</i> , 2009, TRADUÇÃO NOSSA.....	44
QUADRO 8 - CONJUNTO DE VARIÁVEIS E ATRIBUTOS RELACIONADOS À UTILIZAÇÃO.....	50
QUADRO 9 - CONJUNTO DE VARIÁVEIS E DETALHAMENTO DOS ATRIBUTOS RELACIONADOS À UTILIZAÇÃO.....	52
QUADRO 10 - VARIÁVEIS INDEPENDENTES SEGUNDO COUSINS E LEITHWOOD, 1986.....	57
QUADRO 11 - COMPARAÇÃO FATORES QUE INFLUENCIAM A UTILIZAÇÃO. APRESENTADOS POR LEVITON & HUGHES (1981) E COUSINS E LEITHWOOD (1986).	59
QUADRO 12 - RELAÇÃO ENTRE OS FATORES QUE FAVORECEM A UTILIZAÇÃO - IDENTIFICADOS POR COUSINS E LEITHWOOD (1986) E JOHNSON ET AL. (2009).....	67
QUADRO 13 - DESCRIÇÃO DAS CARACTERÍSTICAS (FATORES) DO ENVOLVIMENTO DOS <i>STAKEHOLDERS</i> IDENTIFICADAS POR JOHNSON ET AL. (2009).....	68
QUADRO 14 - MODELO DE MECANISMOS ALTERNATIVOS QUE PODEM MEDIAR A AVALIAÇÃO INFLUENTE.....	75
QUADRO 15 - DISTINÇÕES DE USO.....	77
QUADRO 16 - FATORES QUE FAVORECEM A UTILIZAÇÃO DAS AVALIAÇÕES SEGUNDO LEVITON E HUGHES (1981), COUSINS E LEITHWOOD (1986), JOHNSON (1998) E JOHNSON ET AL. (2009).....	84
QUADRO 17 - RELAÇÃO ENTRE OS PADRÕES DE AVALIAÇÃO DE PROGRAMAS E OS FATORES QUE INFLUENCIAM A UTILIZAÇÃO.....	89
QUADRO 18 - AVALIAÇÕES DE PROGRAMAS SELECIONADAS – TCU.....	99
QUADRO 19 - RELAÇÃO DAS AVALIAÇÕES EXAMINADAS E RESPECTIVOS RESPONDENTES.....	99
QUADRO 20 - PROGRAMAS SELECIONADOS DO PPA 2004-2007 PARA EXAME DAS AVALIAÇÕES.....	100
QUADRO 21 - ESCALA DE AVALIAÇÃO DE OBSERVÂNCIA DOS PADRÕES.....	101
QUADRO 22 - DOCUMENTOS TÉCNICOS DO TCU RELATIVOS À SISTEMÁTICA E ABORDAGEM METODOLÓGICA APLICADAS NAS AUDITORIAS DE NATUREZA OPERACIONAL.....	102
QUADRO 23 - DOCUMENTAÇÃO DAS AVALIAÇÕES DO TCU EXAMINADA POR ESTA PESQUISADORA.....	104

QUADRO 24 - PADRÕES DE AVALIAÇÃO DE PROGRAMAS DEFINIDOS PELO COMITÊ CONJUNTO DE PADRÕES PARA AVALIAÇÃO EDUCACIONAL DOS EUA.....	105
QUADRO 25 - EXTRATO DO QUESTIONÁRIO APLICADO NA PESQUISA PARA COLETA DE DADOS SOBRE A OBSERVÂNCIA DOS PADRÕES DE AVALIAÇÃO DE PROGRAMAS DEFINIDOS PELO COMITÊ CONJUNTO DE PADRÕES PARA AVALIAÇÃO EDUCACIONAL DOS EUA.....	106
QUADRO 26 - TIPOLOGIA DE AVALIAÇÃO GOVERNAMENTAL	115
QUADRO 27 - TIPOLOGIA DAS AVALIAÇÕES REALIZADAS PELO MDS	116
QUADRO 28 - COMPARATIVO DA FREQUÊNCIA DAS AVALIAÇÕES DOS PROGRAMAS NO PERÍODO VIGÊNCIA PPA 2004-2007.....	120
QUADRO 29 - AVALIAÇÕES REALIZADAS PELO TCU NO PERÍODO 2004-2007....	131
QUADRO 30 - SÍNTESE DAS PESQUISAS DE AVALIAÇÃO DE PROGRAMAS SOCIAIS DO MDS	134
QUADRO 31 - QUESTÕES FORMULADAS NA AVALIAÇÃO DOS PROGRAMAS DO PPA 2004-2007 RELATIVAS À AVALIAÇÃO DO PROCESSO	142
QUADRO 32 - RELAÇÃO ENTRE OBJETIVOS E INTERESSADOS NAS AVALIAÇÕES DE PROGRAMAS DO PPA.....	144
QUADRO 33 - TIPOLOGIA DE AVALIAÇÃO DE PROGRAMAS DE GOVERNO NO ÂMBITO DA ADMINISTRAÇÃO PÚBLICA FEDERAL.....	167

LISTA DE SIGLAS

ANSI – American National Standards Institute
ANOp – Auditoria de Natureza Operacional
CGU – Controladoria Geral da União
GAO - General Accountability Office – USA
IPEA - Instituto de Pesquisas Econômicas Aplicadas
JCSEE – Joint Committee on Standards for Educational Evaluation
MAPA – Ministério da Agricultura, Pecuária e Abastecimento
MC – Ministério das Comunicações
MCidades – Ministério das Cidades
MCT – Ministério da Ciência e Tecnologia
MD – Ministério da Defesa
MDA – Ministério do Desenvolvimento Agrário
MDIC – Ministério do Desenvolvimento, Indústria e Comércio Exterior
MDS - Ministério do Desenvolvimento Social e Combate a Fome
ME – Ministério dos Esportes
MEC – Ministério da Educação
MF – Ministério da Fazenda
MJ – Ministério da Justiça
MIN – Ministério da Integração Nacional
MinC – Ministério da Cultura
MMA – Ministério do Meio Ambiente
MME – Ministério de Minas e Energia
MPOG – Ministério do Planejamento Orçamento e Gestão
MPS – Ministério da Previdência Social
MRE – Ministério das Relações Exteriores
MS – Ministério da Saúde
MT – Ministério dos Transportes
MTE – Ministério do Trabalho e Emprego
MTur – Ministério do Turismo
PPA - Plano Plurianual
PR – Presidência da República

SAGI - Secretaria de Avaliação e Gestão da Informação

SEDH – Secretaria Especial de Direitos Humanos

SFC – Secretaria Federal de Controle Interno

SPI – Secretaria de Planejamento e Investimentos Estratégicos

SPM – Secretaria Especial de Políticas para as Mulheres

TCU - Tribunal de Contas da União

TIG – Grupo Temático de Interesse

TMS – Tema de Maior Significância

SUMÁRIO

1 INTRODUÇÃO	15
1.1 CONTEXTUALIZAÇÃO	17
1.2 PROBLEMA	20
1.3 OBJETIVOS	21
1.3.1 Objetivo Geral	21
1.3.2 Objetivos Específicos	22
1.4 DELIMITAÇÃO DO ESTUDO	22
1.5 JUSTIFICATIVA E RELEVÂNCIA DA PESQUISA	23
2 REVISÃO DA LITERATURA E MARCO CONCEITUAL	26
2.1 AVALIAÇÃO DE PROGRAMAS - PROPÓSITOS E TIPOLOGIA	27
2.2 OS PADRÕES DE AVALIAÇÃO DE PROGRAMAS DO JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION	34
2.2.1 Padrões de Utilidade	35
2.2.2 Padrões de Viabilidade	36
2.2.3 Padrões de Propriedade (adequabilidade).....	36
2.2.4 Padrões de Acuracidade.....	38
2.3 DEFINIÇÃO DE UTILIZAÇÃO	39
2.4 DIMENSÕES DA UTILIZAÇÃO	41
2.5 RESGATANDO A DISCUSSÃO SOBRE A UTILIZAÇÃO DA AVALIAÇÃO: O DEBATE WEISS- PATTON (1988).....	45
2.6 A UTILIZAÇÃO DA AVALIAÇÃO: A REVISÃO TEÓRICA E PRÁTICA DE SHULHA E COUSINS (1997)	47
2.7 A PESQUISA SOBRE UTILIZAÇÃO DAS AVALIAÇÕES DE LEVITON E HUGHES (1981)	49
2.7.1 Análise dos fatores de Leviton & Hughes	53
2.8 A PESQUISA SOBRE UTILIZAÇÃO DAS AVALIAÇÕES DE COUSINS E LEITHWOOD (1986)	56
2.8.1 Análise dos fatores de Cousins e Leithwood (1986)	59
2.8.2 Os resultados da análise dos fatores de Cousins e Leithwood (1986).....	61
2.9 A PESQUISA SOBRE A UTILIZAÇÃO DA AVALIAÇÃO - UMA REVISÃO DA LITERATURA EMPÍRICA DE 1986 A 2005 DE JOHNSON <i>ET AL.</i> (2009)	66
2.10 OS MODELOS TEÓRICOS DA UTILIZAÇÃO DA AVALIAÇÃO	68
2.10.1 O meta-modelo de utilização proposto por Johnson (1998).....	69
2.10.2 Avaliações Influentes de Mark e Henry (2004).....	73
2.10.3 O modelo de Patton - Avaliações focadas na utilização.....	75
2.10.4 O mau uso, o abuso, o não uso e o uso não intencional	78
2.11 A NÃO UTILIZAÇÃO DOS RESULTADOS DAS AVALIAÇÕES NO SETOR PÚBLICO	78
2.11.1 Para quem importa os resultados das avaliações?	82
2.11.2 O que dizem as avaliações de programas governamentais?	83
2.12 MARCO CONCEITUAL.....	84
3 METODOLOGIA.....	90
3.1 ESTUDO DE CASOS MÚLTIPLOS	93
3.2 DO DESENVOLVIMENTO DA TEORIA.....	95
3.3 DA SELEÇÃO DOS CASOS	96

3.3.1	– Da Seleção das Avaliações do TCU	98
3.3.2	– Da Seleção das Avaliações de Programas do PPA/SPI/MPOG	99
3.4	ANÁLISE DOS DADOS	100
3.5	COLETA DE DADOS	105
3.6	DAS LIMITAÇÕES DA ESTRATÉGIA METODOLÓGICA	107
4	A AVALIAÇÃO DE PROGRAMAS GOVERNAMENTAIS NO BRASIL	109
4.1	AS AVALIAÇÕES DO PLANO PLURIANUAL 2004-2007 REALIZADAS PELA SPI/MPOG....	117
4.2	AS AVALIAÇÕES DA SECRETARIA FEDERAL DE CONTROLE INTERNO/CGU	120
4.3	AS AVALIAÇÕES DE PROGRAMAS DO TRIBUNAL DE CONTAS DA UNIÃO REALIZADAS NO PERÍODO 2004-2007	125
4.4	AS AVALIAÇÕES REALIZADAS PELA SECRETARIA DE AVALIAÇÃO E GESTÃO DA INFORMAÇÃO - SAGE/MDS	131
5	ANÁLISE DAS AVALIAÇÕES DE PROGRAMAS DO PLANO PLURIANUAL 2004-2007 REALIZADAS PELA SECRETARIA DE PLANEJAMENTO E INVESTIMENTOS ESTRATÉGICOS DO MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO - SPI/MPOG	136
5.1	A ADEQUAÇÃO DAS AVALIAÇÕES DE PROGRAMAS DO PPA 2004-2007 AOS PADRÕES DE AVALIAÇÃO DE PROGRAMAS DO COMITÊ CONJUNTO DE PADRÕES PARA AVALIAÇÃO EDUCACIONAL DOS ESTADOS UNIDOS	139
5.2	A UTILIDADE DA AVALIAÇÃO DE PROGRAMAS DO PPA	146
6	ANÁLISE DAS AVALIAÇÕES DE PROGRAMAS REALIZADAS PELO TRIBUNAL DE CONTAS DA UNIÃO NO PERÍODO 2004-2007	149
6.1	ANÁLISE DA ADEQUAÇÃO DAS AVALIAÇÕES DE PROGRAMAS REALIZADAS PELO TCU AOS PADRÕES DE AVALIAÇÃO DE PROGRAMAS DO COMITÊ CONJUNTO DE PADRÕES PARA AVALIAÇÃO EDUCACIONAL DOS ESTADOS UNIDOS.....	151
7	CONCLUSÕES E RECOMENDAÇÕES.....	165
	REFERÊNCIAS	174
	APÊNDICE	184

1 INTRODUÇÃO

Esta pesquisa trata do estudo da avaliação de programas governamentais sob a perspectiva da utilização dos seus resultados. A utilização tem sido um campo muito explorado internacionalmente, cujas pesquisas iniciaram no final dos anos 70, quando os avaliadores passaram a questionar a efetiva utilidade das avaliações para influenciar decisões e contribuir para o aperfeiçoamento de programas e políticas.

Como no Brasil os estudos ainda são incipientes e com vistas a oferecer uma contribuição para a compreensão deste tema, – principalmente, quanto aos seus aspectos conceituais e práticos – este trabalho apresenta uma análise empírica dessas pesquisas aplicadas ao caso brasileiro, além de uma consolidação da literatura relativa às pesquisas realizadas, com destaque para aquelas que fundamentam a trajetória evolutiva do conhecimento atual sobre a utilização da avaliação.

Assim, a partir da revisão bibliográfica, foram identificados os fatores que influenciam a utilização das avaliações, os quais foram examinados à luz dos padrões de avaliação de programas definidos pelo *Joint Committee on Standards for Educational Evaluation*, organização formada pela associação das principais entidades profissionais responsáveis pela atividade avaliativa em educação nos Estados Unidos, cujos padrões são considerados Padrões Nacionais Americanos e orientam a prática dessa atividade naquele país, sendo amplamente referendados pela literatura pesquisada, como mecanismos para assegurar a qualidade e credibilidade das avaliações.

O objetivo dessa estratégia de pesquisa foi identificar os critérios a serem utilizados para examinar as avaliações de programas governamentais no Brasil, com vistas a analisar em que medida as sistemáticas e abordagens metodológicas aplicadas favorecem a utilização dos resultados por elas produzidos, principalmente, no que diz respeito àquelas realizadas pelas organizações públicas responsáveis pela função avaliativa.

Desta forma, por meio de levantamento documental na legislação pertinente à administração pública federal, foram identificados os tipos de avaliações existentes, seus objetivos, características e a quem se destinam, com vistas a elaborar o perfil das avaliações governamentais e classificá-las de acordo com os propósitos elencados na literatura revisada.

De posse dessas informações, procedeu-se ao estudo das sistemáticas e abordagens metodológicas utilizadas para realização das avaliações dos programas do Plano Plurianual 2004-2007, sob a coordenação da Secretaria de Planejamento e Investimentos Estratégicos do Ministério do Planejamento, Orçamento e Gestão; bem como das avaliações de programas do

Tribunal de Contas da União, realizadas pela Secretaria de Fiscalização e Avaliação de Programas Governamentais, por intermédio das auditorias operacionais; das avaliações incumbidas à Secretaria Federal de Controle Interno da Controladoria Geral da União, processadas por meio de fiscalizações e auditorias; e ainda, das avaliações decorrentes da Política de Monitoramento e Avaliação do Ministério do Desenvolvimento Social e Combate à Fome, sob a responsabilidade da Secretaria de Avaliação e Gestão da Informação.

Para a realização da pesquisa empírica, utilizou-se a abordagem metodológica de estudos de casos múltiplos, com base em Yin (2005). O objetivo foi o de examinar em que medida as estratégias adotadas para realização das avaliações mencionadas observaram os padrões de avaliação definidos pelo *Joint Committee*, e, em decorrência, como e em que medida foram implementadas, considerando os fatores que influenciam a utilização, na extensão em que incorporam, nas respectivas abordagens, as necessidades e prioridades daqueles atores e/ou organizações que deveriam utilizar os resultados e informações produzidos pelas avaliações para dar cumprimento aos objetivos propostos.

Cabe destacar, no entanto, que alguns fatores comprometeram o pleno cumprimento dos objetivos desta pesquisa, principalmente no que diz respeito à parte empírica, em razão da limitação de prazo para conclusão dos trabalhos, bem como da não disponibilização tempestiva das informações, fatos que impossibilitaram contemplar a análise dos dados das avaliações realizadas pela Secretaria Federal de Controle Interno – SFC/CGU e da Secretaria de Avaliação e Gestão da Informação – SAGI/MDS, muito embora as respectivas sistemáticas de avaliação tenham sido levantadas e apresentadas no capítulo que trata da institucionalização da função avaliação no Brasil.

Em que pese essas limitações, foram coletados e analisados os dados e informações de uma amostra não estatística das avaliações dos programas do PPA 2004-2007 realizadas pela Secretaria de Planejamento e Investimentos Estratégicos – SPI/MPOG e das avaliações executadas pela Secretaria de Fiscalização e Avaliação de Programas do Tribunal de Contas da União no mesmo período, favorecendo, assim, a caracterização dessas avaliações em relação a sua adequabilidade aos padrões de avaliação, sem a pretensão de se generalizar os achados observados, apenas com o intuito de evidenciar o nível de aderência à prática avaliativa, com ênfase na utilização, com base nos padrões mencionados.

Para tanto, este estudo foi desenvolvido ao longo de 7 capítulos, sendo que, nesta introdução, contextualiza-se o tema da pesquisa, o problema e as questões a serem investigadas. O capítulo 2 trata da revisão de literatura e marco conceitual, considerando os propósitos e a tipologia de avaliação de programas, a importância dos padrões, as pesquisas e

teorias desenvolvidas sobre os fatores que influenciam a utilização, bem como sua relação com os padrões de avaliação de programas definidos pelo *Joint Committee* e, ainda, algumas evidências sobre a não utilização das avaliações no setor público. O capítulo 3 trata da metodologia utilizada; o capítulo 4 da institucionalização da função avaliação no Brasil, com base na qual são identificados os tipos de avaliações existentes no âmbito da administração pública federal, suas características, objetivos e forma de implementação; e, as análises dos dados estão apresentadas nos capítulos 5 e 6. Por fim, o capítulo 7 apresenta as conclusões e recomendações da pesquisa realizada.

1.1 CONTEXTUALIZAÇÃO

As avaliações de programas governamentais no Brasil têm merecido crescente interesse nas últimas décadas, tanto no meio acadêmico como no âmbito da administração pública. Pode-se notar esse fato, em especial, a partir da Reforma do Aparelho do Estado de 1995, onde passou a ocupar papel estratégico no ciclo de gestão, como instrumento para subsidiar o controle dos resultados, objetivando a melhoria da eficiência e da qualidade na prestação dos serviços públicos, bem como no desenvolvimento de uma cultura gerencial nas organizações (BRASIL, 1995).

A institucionalização da função avaliação está associada à busca do aperfeiçoamento de programas e políticas públicas, com vistas a assegurar maior efetividade e transparência das ações governamentais, mediante a melhoria da capacidade de gestão, por intermédio de informações e análises que favoreçam as melhores escolhas alocativas, de forma a propiciar o alcance mais eficaz e eficiente dos resultados dessas ações para promoção da melhoria social (SILVA; COSTA, 2002).

Para dar cumprimento a esses propósitos, o sistema governamental de planejamento e orçamento foi reestruturado a partir do Plano Plurianual 2000-2003, adotando-se o modelo gerencial de orçamento e gestão orientado para resultados, no qual todas as despesas foram organizadas em programas voltados à solução de problemas e atendimento a demandas da sociedade.

Para o controle dos resultados, foi criado um sistema de monitoramento e avaliação, que anualmente disponibiliza, ao Congresso e à sociedade, relatórios contendo análise do desempenho dos programas governamentais.

Essas mudanças impuseram, de igual modo, transformações nas organizações públicas responsáveis pelo cumprimento das novas demandas impostas pelo Estado, tanto nas suas estruturas, como nos seus processos e instrumentos de trabalho.

Nesse sentido, foram organizados os sistemas de planejamento e de orçamento, de administração financeira e de contabilidade federal, bem como de controle interno do Poder Executivo Federal. Assim, coube ao Ministério do Planejamento, Orçamento e Gestão - MPOG, em especial à Secretaria de Planejamento e Investimentos Estratégicos - SPI, como também às demais unidades setoriais responsáveis pelas atividades de planejamento, a competência, dentre outras, de avaliar os planos e programas de governo, quanto à eficácia e efetividade, com vistas a subsidiar o processo de alocação de recursos públicos, a política de gastos e a coordenação; e ainda, assegurar que as unidades responsáveis pela execução dos programas, projetos e atividades da administração federal mantivessem rotinas de acompanhamento e avaliação da sua programação (BRASIL, 2001a).

Ao sistema de controle interno do Poder Executivo, por intermédio da Secretaria Federal de Controle Interno - SFC da Controladoria Geral da União - CGU e das Secretarias de Controle Interno dos Poderes Legislativo e Judiciário, coube a avaliação da ação governamental e da gestão dos administradores públicos federais, por intermédio da fiscalização contábil, financeira, orçamentária, operacional e patrimonial (BRASIL, 2001a).

Cabe observar, no entanto, que o Brasil adota o sistema duplo de controle da administração pública Federal: o controle externo, de competência do Congresso Nacional, com o auxílio do Tribunal de Contas da União - TCU; e o controle interno, exercido pelo sistema de controle interno de cada Poder (BRASIL, 1988).

Compete ao TCU, no exercício do controle externo, normatizar e realizar inspeções e auditorias de natureza contábil, financeira, orçamentária, operacional e patrimonial, com vistas a verificar a legalidade, a economicidade, a legitimidade, a eficiência, a eficácia e a efetividade de atos, contratos e fatos administrativos praticados no âmbito da administração pública federal (BRASIL, 1992 e 2002).

As reformas gerenciais do Estado, nas últimas décadas, favoreceram a ampliação das competências dos órgãos de auditoria governamental, conforme observado por Barzelay (1997), introduzindo novas modalidades de auditoria, dentre elas a auditoria de desempenho (*performance auditing*), que, rapidamente, foi aceita como um tipo de revisão da atividade governamental *a posteriori*, voltada para a aferição dos resultados dessas ações, cujo principal objetivo pode ser expresso como *performance accountability*, que se relaciona à responsabilização das organizações pelos resultados de suas atividades.

Barzelay (1997) constatou ainda que a auditoria de desempenho apresenta características similares à avaliação de programas, levando-o inclusive a concluir que essa modalidade de auditoria é, na realidade, avaliação de programa, porque envolve a produção e revisão de instrumental de julgamento.

No Brasil, a modalidade de auditoria voltada à aferição do desempenho e dos resultados da ação governamental denomina-se auditoria operacional, conforme expresso na Constituição Federal e definido nos normativos do Tribunal de Contas da União (BRASIL, 1988, 1992, 2000 e 2002).

Pode-se depreender, em síntese, que a função avaliação foi instituída no país tanto como instrumento para subsidiar e retroalimentar as funções de planejamento, orçamento e gestão governamentais, como também para o controle da administração pública, que se realiza por meio das fiscalizações e auditorias, tanto no âmbito do sistema de controle interno, como do controle externo por intermédio do TCU.

No entanto, em que pese a institucionalização formal da função avaliação no Brasil, com a produção de inúmeros estudos pelas organizações públicas responsáveis por essa atividade, pouca atenção tem sido dada à utilização das informações e resultados das avaliações de programas governamentais. Trata-se de uma questão essencial e que tem gerado um amplo questionamento sobre as práticas de avaliação adotadas em âmbito internacional, assim como sobre a forma como a função avaliação tem se estruturado, principalmente, nas democracias avançadas (WEISS, 1988a; RIST, 1990, CHELIMSKY, 2009).

No Brasil, esse debate é incipiente. Aqui ainda prevalece disseminada a crença da existência de uma relação simples e linear entre avaliação e ação (utilização instrumental), como aquela preconizada no modelo racional de tomada de decisão (March, 1994). Nessa perspectiva, supõe-se que, uma vez disponibilizadas, as avaliações e o conhecimento por elas produzidos serão utilizados para atender às demandas de múltiplos atores. Prevalece a percepção de que a avaliação é capaz de fornecer informações objetivas, livres de julgamento de valor e que atenderiam a todos os propósitos.

No entanto, alguns trabalhos evidenciam a ausência de estudos empíricos que permitam mensurar o grau de utilização das avaliações (CALMON E GUSSO, 2003). Dessa forma, para compreensão desse fenômeno, torna-se necessário investigar, a partir da institucionalização da função avaliação no Brasil, como os sistemas avaliativos governamentais foram constituídos, quais os tipos de avaliação existentes, quais são seus objetivos, características e para quem se destinam, bem como identificar e mensurar com base nos padrões definidos pelo *Joint Committee on Standards for Educational Evaluation*, em que

medida as avaliações governamentais foram concebidas de forma a atender as necessidades de informações daqueles que farão uso dos seus resultados e informações.

1.2 PROBLEMA

Segundo Weiss (1988 e 1993) e Radin (2006), os estudos realizados sobre a utilização das avaliações demonstram que as informações e os resultados por elas produzidos são pouco utilizados para promoção de ações aperfeiçoadoras no âmbito de programas e políticas, pelas seguintes razões:

- i) as avaliações são realizadas com base na premissa de que os resultados serão, direta e automaticamente, utilizados para promover as melhorias necessárias, sem que sejam consideradas previamente as necessidades de utilização dos atores que deveriam fazer uso das informações e dos resultados por elas fornecidos;
- ii) as avaliações governamentais geralmente contemplam múltiplos objetivos, ou seja, atender tanto às demandas para a melhoria da gestão, dos programas, das organizações, da alocação de recursos, das políticas; assim como promover a *accountability* governamental e o controle social;
- iii) as abordagens metodológicas utilizadas nem sempre são adequadas para favorecer o alcance de todos os objetivos propostos;
- iv) os mecanismos para mensuração da utilização dos resultados das avaliações não são adequados para refletir a efetiva contribuição das avaliações para promoção dos benefícios esperados.

Com base nessas considerações, o problema desta pesquisa foi definido nos seguintes termos:

– Para que (objetivos), para quem (usuários) e de que forma (métodos) são realizadas as avaliações de programas governamentais, especialmente aquelas sob a responsabilidade da SPI/MPOG, SFC/CGU e TCU, e em que medida as estratégias e abordagens adotadas por esses órgãos favorecem a utilização dos resultados com vistas ao alcance dos objetivos propostos?

1.3 OBJETIVOS

As avaliações de programas realizadas pelas organizações da administração pública federal, em especial, pela SPI, SFC e TCU têm por objetivo contribuir para o aperfeiçoamento de programas e políticas, para a melhoria da gestão e da alocação de recursos, assim como para promoção do aprendizado organizacional e da *accountability* governamental. No entanto, alguns trabalhos evidenciam que, no Brasil, não obstante a produção e disponibilização sistemática dessas avaliações, não têm sido identificados estudos empíricos sobre a utilização dos seus resultados, que permitam conhecer em que medida e de que forma as avaliações contribuem para dar cumprimento aos objetivos mencionados.

Para analisar essa questão, no entanto, é necessário examinar se as estratégias e abordagens metodológicas adotadas pelas entidades citadas favorecem a identificação tanto dos atores e organizações responsáveis em dar cumprimento aos objetivos, como das suas necessidades de informações; tendo em vista que esses elementos são fundamentais para a definição do objeto e do objetivo da avaliação governamental, e determinam a forma como será conduzida, ou seja, a abordagem metodológica a ser empregada.

Desta forma, o problema apresentado, expresso no objetivo geral a seguir descrito, será examinado por intermédio dos objetivos específicos relacionados no item 1.3.2.

1.3.1 Objetivo Geral

Este trabalho tem por objetivo, a partir da pesquisa bibliográfica sobre a utilização das avaliações e do levantamento documental, identificar e examinar os propósitos, a tipologia, os objetivos, as sistemáticas e metodologias empregadas, a quem se destinam e como são utilizados os resultados das avaliações de programas governamentais no Brasil, especialmente aquelas realizadas pela SPI/MPOG, SFC/CGU e TCU, com vistas a verificar em que medida as estratégias e abordagens metodológicas adotadas consideram as necessidades e prioridades daqueles atores e/ou organizações que devem utilizar os resultados e informações produzidos pelas avaliações para dar cumprimento aos objetivos propostos.

Para tanto, por meio de estudos de casos múltiplos, serão aplicados os padrões de avaliação de programas definidos pelo *Joint Committee on Standards for Educational Evaluation*: utilidade, viabilidade, propriedade e acuracidade, os quais, segundo Sanders (1994), se observados, são suficientes para produzir avaliações confiáveis e imparciais (SANDERS, 1994, p. 5).

1.3.2 Objetivos Específicos

Os objetivos específicos são os seguintes:

- i) realizar pesquisa bibliográfica com vistas a identificar os principais aspectos conceituais a serem aplicados no estudo da utilização das avaliações de programas governamentais;
- ii) identificar os tipos, objetivos, características e para quem se destinam as avaliações de programas governamentais existentes no âmbito da administração pública federal no Brasil;
- iii) examinar a credibilidade e qualidade das avaliações realizadas pela SPI, SFC e TCU com base nos padrões de avaliação de programas definidos pelo *Joint Committee for Educational Evaluation*, com ênfase na utilização;
- iv) definir o perfil das avaliações governamentais realizadas pela SPI, SFC e TCU, considerando os objetivos, metodologia e os usuários dos resultados produzidos pelas avaliações, com vistas a indicar a adequação e consistência das abordagens adotadas às necessidades daqueles que farão uso dessas informações;
- v) identificar os fatores e os mecanismos que influenciam a utilização dos resultados das avaliações com base na revisão da literatura;
- vi) examinar a relação entre os padrões de avaliação e os fatores e mecanismos que influenciam a utilização das avaliações, com vistas a identificar em que medida as abordagens adotadas nas avaliações realizadas pela SPI, SFC e TCU favorecem a sua utilização.

1.4 DELIMITAÇÃO DO ESTUDO

Esta pesquisa tem por objeto as avaliações de programas governamentais realizadas no Brasil. Desta forma, é importante ressaltar que, no âmbito da administração pública federal, *programa* é definido como o instrumento de organização das ações de governo, com vistas à elaboração dos orçamentos públicos. Portanto, as avaliações a serem consideradas neste estudo dizem respeito apenas aos programas constantes do Plano Plurianual. Assim, para realização dos estudos de casos múltiplos, em que pese tenha sido inicialmente proposto examinar as avaliações de programas realizadas pela SPI, SFC e TCU, foram analisadas em profundidade, apenas, as avaliações de programas do PPA 2004-2007, sob a responsabilidade

da SPI/MPOG e as avaliações realizadas pelo Tribunal de Contas da União, no mesmo período.

Essa limitação decorreu da dificuldade em se obter, no prazo estabelecido para conclusão deste trabalho, os dados necessários para o desenvolvimento da pesquisa, relativos às fiscalizações e auditorias realizadas pela SFC/CGU, instrumentos por intermédio dos quais são avaliados os programas de governo.

1.5 JUSTIFICATIVA E RELEVÂNCIA DA PESQUISA

Conforme apresentado na contextualização do tema, a combinação do modelo racional de tomada de decisão com a perspectiva positivista da filosofia da ciência tem resultado em um receituário para as avaliações realizadas no país. Esse receituário se fundamenta em um sistema de crenças sobre o *design* e o papel a ser desempenhado pela função avaliação no governo brasileiro, que demanda cada vez mais avaliações que não levam devidamente em conta a questão da utilização.

Desta forma, o estudo sobre a utilização das avaliações de programas governamentais no Brasil mostra-se relevante e oportuno, ante a sua potencial contribuição para o desenvolvimento de modelos avaliativos que assegurem maior efetividade dos benefícios esperados com a realização desses trabalhos, contribuindo para o fortalecimento e institucionalização da função avaliação no país.

Ademais, a literatura disponível sobre este assunto, em português, ainda não propicia a disseminação dos conceitos e das teorias já desenvolvidas em outros países. Desta forma, mediante a apresentação da revisão da literatura no idioma pátrio é oferecida uma consolidação das pesquisas realizadas, com destaque para os mais importantes estudos realizados sobre esse tema, de forma a favorecer a compreensão da estrutura conceitual do objeto pesquisado.

Além disso, a abordagem da avaliação sob a perspectiva da utilização é um tema que possibilitará a inserção brasileira nas discussões internacionais, de forma a contribuir, a partir da pesquisa empírica, para o desenvolvimento das teorias da utilização.

As teorias da utilização buscam investigar e explicar como as avaliações são utilizadas; que fatores influenciam a utilização dos seus resultados; quais os mecanismos para mensurar a utilização; e, quais as características metodológicas dos estudos empíricos para investigar o uso da avaliação.

Segundo Patton (2008):

[...] a questão da utilização da avaliação surgiu da interface entre ciência e ação, entre conhecer e fazer. Ela traz à discussão questões fundamentais sobre a racionalidade humana, sobre o processo decisório e o conhecimento aplicado para construção de um mundo melhor. (PATTON, 2008, p. 5, tradução nossa).

Para Shulha e Cousins (1997) a questão da utilização é uma preocupação antiga e talvez a maior fonte de frustração entre os avaliadores, já Mark, Henry e Julnes (2000, p. 21) consideram que a literatura sobre utilização parece ter surgido, pelo menos em parte, quando os avaliadores reconheceram que os resultados das avaliações não estavam sendo traduzidos em decisões imediatas e observáveis, diretamente relacionadas aos achados nelas reportados.

Moynihan (2008), por exemplo, ao apresentar dez razões para se repensar a gestão por resultados, ressaltou que o principal desafio é promover o uso das informações disponibilizadas pelos sistemas de monitoramento e avaliação, cada vez mais disseminados como instrumentos garantidores da transparência e da *accountability* dos gastos públicos.

No entanto, para serem utilizáveis, as avaliações precisam propiciar aos *stakeholders* o conhecimento necessário sobre o desempenho do programa, tanto quanto dos seus aspectos positivos, como negativos, para os quais deveriam ser formuladas recomendações. As avaliações de programa, em regra, deveriam contribuir para facilitar a escolha de decisões apropriadas, pelos *stakeholders*¹, para a solução dos problemas que comprometem ou dificultam o alcance dos objetivos dos programas por eles geridos (CRONBACH, 1982 *in* CHEN *apud* ALKIN, 2004).

Ocorre que, conforme sintetizado por Shulha e Cousins (1997), a relação avaliador-avaliado e o produto dessa relação – a avaliação e seus resultados – estão inseridos em um contexto multidimensional, cujos fatores epistemológicos, políticos e organizacionais interconectam-se produzindo efeitos sobre a sua utilização.

Esses efeitos, por outro lado, afetam e são afetados pelos atores envolvidos no processo de avaliação, tanto individualmente, por meio das percepções, preferências, interesses, valores, comportamento, como pela inter-relação entre esses atores, as organizações, suas estruturas, processos e ambientes, e ainda outros fatores, como o tempo e o contexto em que estão inseridas (KIRKART, 2000, MARK e HENRY, 2004, PRESKILL e CARACELLI, 1997, TORRES e PRESKILL, 2001).

¹ Pessoas envolvidas, interessadas ou que podem vir a ser afetadas pela avaliação.

No Brasil, em que pese a disseminação da função avaliativa, principalmente após a implementação do modelo gerencial de atuação do Estado, pouco tem sido estudado sobre a efetividade da avaliação, como instrumento para promoção do aperfeiçoamento da gestão governamental, dos programas e das políticas públicas.

Assim, o objetivo desse trabalho é contribuir para a compreensão dos fatores que influenciam a utilização das avaliações e examinar em que medida as avaliações de programas governamentais, no Brasil, foram concebidas e implementadas de forma a favorecer a utilização dos seus resultados, com vistas a oferecer subsídios para se repensar avaliação governamental, em termos do cumprimento dos seus propósitos, buscando evidenciar readequar as abordagens metodológicas a essas finalidades.

2 REVISÃO DA LITERATURA E MARCO CONCEITUAL

Neste capítulo são apresentados, inicialmente, os propósitos e a tipologia da avaliação, com vistas a introduzir os padrões de avaliação de programas definidos pelo *Joint Committee on Standards for Educational Evaluation* e demonstrar a sua relevância para examinar a adequação dos propósitos à utilização das avaliações. Em seguida, apresenta-se a definição e as dimensões da utilização para melhor compreensão do fenômeno em estudo.

Para contextualizar o tema desta pesquisa na literatura, iniciamos com o debate entre dois proeminentes teóricos, Carol Weiss e Michael Patton, que pode ser considerado um marco na construção das teorias que se seguiram, por que trouxe à discussão diferentes visões, valores e perspectivas em relação aos propósitos da avaliação, demonstrando que esse instrumento serve a uma multiplicidade de funções. Em seguida, apresentamos a revisão da teoria, pesquisa e prática da utilização da avaliação, elaborada por Shulha e Cousins (1997), que reforça a contribuição do debate Weiss-Patton e evidencia os novos rumos que se tem tomado no estudo da utilização.

Após a revisão de Shulha e Cousins, iniciamos o estudo da utilização, propriamente dito, buscando seguir uma ordem cronológica em relação às pesquisas desenvolvidas. Apresenta-se, a princípio, o trabalho de Leviton e Hughes (1981), que classifica os fatores que influenciam a utilização dos resultados, seguido da pesquisa de Cousins e Leithwood (1986). Nesse estudo, os autores calcularam um índice das relações entre a utilização (variável dependente) e um conjunto de fatores (variáveis independentes) relacionados a duas categorias (contextos da implementação da avaliação e da decisão/política).

Na sequência dessas pesquisas, é apresentado o estudo desenvolvido por Johnson *et al.* (2009), uma revisão da literatura empírica sobre a utilização da avaliação, compreendendo o período de 1986 a 2005, na qual foi utilizada a estrutura do trabalho de Cousins e Leithwood (1986) para categorizar os estudos sobre a utilização da avaliação. Como conclusão do trabalho, os autores identificaram uma nova categoria e um novo fator relacionado ao contexto da implementação da avaliação, que influenciam a utilização. São eles: o envolvimento dos *stakeholders* (categoria) e a competência do avaliador (fator).

Com vistas a demonstrar a aplicação dos fatores que influenciam a utilização, são apresentados três modelos teóricos da utilização da avaliação: o meta-modelo de Johnson (1998), as avaliações influentes de Mark e Henry (2004) e a teoria revisada da avaliação focada na utilizada de Patton (2008).

Para corroborar a importância do tema desta pesquisa, inserimos, ainda, algumas evidências da literatura sobre gestão pública, que reforçam os argumentos de que as avaliações de programas governamentais, da forma como são concebidas e realizadas, pouco favorecem a utilização dos resultados.

Finalizando este capítulo, apresenta-se a relação entre os padrões de avaliação de programas definidos pelo *Joint Committee on Standards for Educational Evaluation* (Comitê Conjunto de Padrões para Avaliação Educacional) e os fatores que influenciam a utilização – apresentados nas seções anteriores – com vistas a explicitar o marco conceitual que embasou a operacionalização deste estudo, ratificando a importância dos padrões como mecanismos que propiciam a utilização das avaliações, para dar cumprimento aos propósitos que lhes deram origem.

2.1 AVALIAÇÃO DE PROGRAMAS - PROPÓSITOS E TIPOLOGIA

A avaliação, segundo Weiss (1998), é um termo bastante abrangente que acomoda muitas definições. No entanto, o que todas elas têm em comum é a noção de julgamento de mérito, baseado em critérios, segundo um método específico. Para Weiss a avaliação é “uma análise sistemática do processo e/ou dos resultados de um programa ou política, em comparação com um conjunto explícito ou implícito de padrões, com o objetivo de contribuir para o seu aperfeiçoamento” (WEISS, 1998, p. 4, tradução nossa).

Essa definição pode ser decomposta em cinco elementos: o primeiro diz respeito à análise sistemática, onde se enfatiza a natureza metodológica da pesquisa; o segundo e o terceiro referem-se ao foco da investigação, ou seja, o processo de execução ou os resultados do programa e/ou política; o quarto trata dos critérios, que são os padrões de comparação; e o quinto relaciona-se ao propósito da avaliação: contribuir para a melhoria do programa e/ou política, por meio da utilização dos resultados por pessoas com autoridade para tomar decisões (WEISS, 1998).

Com base no último elemento apresentado por Weiss, pode-se depreender que o propósito da avaliação pressupõe a existência de duas variáveis: o objetivo do trabalho (para quê?) e o usuário dos resultados (para quem?).

Segundo os padrões de avaliação estabelecidos pelo *Joint Committee on Standards for Educational Evaluation*, os propósitos de uma avaliação são definidos em termos dos seus objetivos e na intenção de utilização dos seus resultados (SANDERS, 1994, p. 137).

Exemplificando, Sanders (1994) assim apresenta:

[...] o objetivo de uma avaliação em particular pode ser o julgamento de mérito de livros-textos para um curso de curta duração, um treinamento de três semanas; a utilização correspondente que se espera, é que os achados ajudem os instrutores a escolher o livro-texto (SANDERS, 1994, p. 137, tradução nossa).

Segundo Chelimsky (1997), “o propósito da avaliação condiciona a utilização que pode ser dela esperada” (CHELIMSKY, 1997 apud PATTON, 2008, p. tradução nossa).

Weiss (1988) classifica os propósitos das avaliações em duas categorias: para influenciar a tomada de decisão ou como aprendizado. As avaliações que têm por finalidade a tomada de decisão podem ter por objetivos:

- i) corrigir o curso de um programa ou política com vistas a obter melhorias nos processos de execução;
- ii) verificar o alcance dos objetivos com vistas à continuação, expansão ou institucionalização de um programa-piloto; assim como corte, finalização ou abandono de programa em execução;
- iii) testar novas ideias sobre determinado programa;
- iv) escolher a melhor alternativa, dentre várias opções, para implementar ações/medidas; e
- v) decidir sobre a continuidade de um financiamento (WEISS, 1998).

As avaliações que têm como propósito o aprendizado podem ter como objetivos:

- i) documentar a história do programa de forma a contribuir para a sua compreensão e influenciar decisões futuras;
- ii) contribuir como *feedback* para os executores disponibilizando informações sobre o andamento de determinado programa;
- iii) influenciar o comportamento dos executores por intermédio da ênfase nos objetivos do programa;
- iv) promover a *accountability*;
- v) compreender a intervenção social (WEISS, 1998).

No que diz respeito aos interessados na utilização das avaliações, Weiss (1998) faz uma distinção entre os interesses dos políticos e altos executivos, e, o interesse dos gerentes. Segundo a autora, os políticos e altos executivos estão mais preocupados com a efetividade do programa, sendo que os primeiros têm maior interesse nos efeitos do programa, nos seus

impactos; e os últimos, na sua estratégia de implementação, no que deve ser feito. Já os gerentes estão preocupados com as ações do dia a dia, com o operacional (como fazer). Além desses atores, a autora ainda destaca outros interessados, como os financiadores, o público (contribuintes), os usuários dos serviços, os gerentes de programas similares, os formuladores de política e programas e os cientistas sociais (WEISS, 1998).

Desta forma, com todas as possibilidades de interesses e interessados que a avaliação pode servir, o avaliador tem de fazer escolhas. É impossível atender a todos os propósitos dentro dos limites de um único estudo. Muito embora seja possível contemplar diferentes tipos de questões no mesmo trabalho, nem mesmo um estudo bem planejado poderá fornecer informações para todas as questões apresentadas por todos os atores envolvidos. Há que se considerar, ainda, a possibilidade de incompatibilidade entre os propósitos (WEISS, 1988, p.33).

Cabe observar, que diferentes propósitos relacionam-se a diferentes formas de utilização e têm implicações no desenho da avaliação; na mensuração, análise e interpretação dos dados; bem como na forma de apresentação do relatório, na disseminação dos resultados e nos critérios para julgamento da qualidade do trabalho (PATTON, 2008).

Para reduzir a complexidade que caracteriza o campo da avaliação, Patton (2008) propõe seis alternativas de abordagens categorizadas de acordo com os propósitos da avaliação, com vistas a atender, primeiramente, os interessados na utilização.

A primeira abordagem apresentada por Patton (2008) são as avaliações somativas, aquelas orientadas para o julgamento, cujo objetivo é determinar o mérito em geral, valor ou significância de alguma coisa. Patton faz uma distinção entre mérito e valor. O mérito é intrínseco ao programa, é examinado a partir da perspectiva daqueles que são beneficiados por um determinado programa. Por outro lado, valor é extrínseco, diz respeito aos efeitos do programa para aqueles que não são seus beneficiários diretos, como a sociedade em geral.

As avaliações somativas são utilizadas para subsidiar decisões, como, por exemplo, continuar, expandir ou modificar um programa. Nesse tipo de avaliação, a especificação dos critérios para julgamento é central e crítica. Identificar os valores que servirão de base para o julgamento é um dos principais papéis do avaliador, e constitui um dos padrões da utilização definidos pelo *Joint Committee on Standards for Educational Evaluation*.

Em seguida, Patton trata das avaliações formativas que são orientadas para o aperfeiçoamento e têm por objetivo promover melhorias por intermédio da coleta de uma variedade de informações sobre forças e fraquezas do programa e que serão utilizadas para subsidiar um processo de reflexão e inovação em andamento.

O terceiro tipo diz respeito às avaliações que têm por propósito a *accountability*. São aquelas cujo objetivo é o exame da conformidade e a identificação de falhas, ou seja, “em que medida um programa observa as diretrizes, regulamentos, padrões legais ou qualquer outra expectativa formal”. (MARK, HENRY E JULNES, 2000, p. 13, tradução nossa). Uma abordagem comumente utilizada nesse tipo de avaliação é a análise de indicadores de desempenho. No entanto, Patton ressalta que, segundo Burt Perrin² (2002, 1998), a utilização de indicadores para avaliação é limitada. Os indicadores sozinhos não esclarecem as causas dos resultados, dificultando a tomada de decisão ou a solução de problemas. Os sistemas de *accountability* servem, principalmente, ao propósito de prestar contas, tanto no setor público como privado (PERRIN, 2002, 1998 apud PATTON, 2008).

Para Patton (2008), no setor público, as auditorias de programa e financeiras destinam-se a examinar a conformidade dos programas governamentais com os propósitos e procedimentos legais. Nos Estados Unidos, as unidades de avaliação de programas das agências de auditoria do legislativo, controladorias e outras afins têm a responsabilidade de assegurar tanto que os programas estão sendo adequadamente implementados, como também, que são efetivos. Para tanto, examinam questões que dizem respeito, por exemplo, a: se os recursos estão sendo utilizados para atender aos propósitos estabelecidos; se os objetivos e as metas estão sendo alcançados; se os indicadores apresentam melhorias; se os recursos estão sendo eficientemente alocados; se os problemas estão sendo controlados; se o pessoal é qualificado; se somente o público elegível está sendo aceito no programa; se a implementação está observando o plano aprovado; e, se existem e são utilizados mecanismos de controle de qualidade (PATTON, 2008, p. 121-122).

A quarta abordagem apresentada por Patton é o monitoramento. “Algumas vezes o monitoramento é subentendido como uma modalidade da *accountability*, porque ambos utilizam indicadores de desempenho.” (PATTON, 2008, p. 126, tradução nossa). No entanto, os indicadores de desempenho podem servir para diferentes propósitos, tanto para *accountability*, quanto para o acompanhamento da gestão. Neste caso, os propósitos envolvem diferentes interessados primários na utilização, o que pode gerar conflitos. A abordagem de avaliação que tem como propósito a *accountability* é direcionada para os *stakeholders* externos e para aqueles que são responsáveis pelo programa. Já o monitoramento atende às

² PERRIN, Burt. Effective Use and Misuse of Performance Measurement. *American Journal of Evaluation*, 19, p. 367-379, 1998.

_____. Towards a New View of Accountability. Paper presented to the European Evaluation Society annual conference, Seville, Spain, October, 11, 2002.

necessidades internas, mediante a disponibilização de informações sobre o programa, para melhorar o seu gerenciamento.

Patton ressalta a importância em se distinguir o monitoramento como uma abordagem da avaliação, porque nos países em desenvolvimento tem-se desenvolvido sistemas de monitoramento e avaliação (M&E) que tem como propósito tanto a *accountability* quanto o gerenciamento. Isso pode dificultar a utilização das informações por eles produzidas, porque seus objetivos são diferentes, assim como os interessados na sua utilização. Em geral os políticos e financiadores necessitam de informações agregadas e os gerentes, de informações detalhadas. Conciliar indicadores que contemplem ambas as necessidades é o grande desafio dos sistemas de monitoramento de desempenho (PATTON, 2008).

No que diz respeito à quinta abordagem proposta por Patton (2008) – avaliações para produção de conhecimento – tem-se, como unidade de análise, os achados de diferentes avaliações de programas, que são examinados com o objetivo de identificar e formular generalizações sobre os processos e intervenções que contribuíram para a efetividade dos programas estudados.

O último dos propósitos apresentados por Patton trata da abordagem da avaliação voltada para os ambientes dinâmicos, cujas condições exigem adaptações constantes. São as avaliações desenvolvimentistas. Ao contrário das avaliações formativas, orientadas para o aperfeiçoamento, que têm como foco melhorias nas intervenções ou nos modelos, as avaliações desenvolvimentistas direcionam-se para mudanças na intervenção, adaptando-as a novas circunstâncias, alterando as táticas com base nas condições emergentes.

Mark, Henry e Julnes (2000), por outro lado, identificaram quatro propósitos primários, para os quais os resultados das avaliações podem ser utilizados, que parecem contemplar os propósitos apresentados tanto por Weiss, como por Patton. São eles: julgamento do mérito e valor, aperfeiçoamento de programas e organizações, conformidade e identificação de falhas, e, por último, desenvolvimento de conhecimento.

Para esses autores, as avaliações existem para contribuir para o aperfeiçoamento de programas e políticas por intermédio da disponibilização de informações que podem ser utilizadas pelas instituições democráticas para promoção da melhoria social (MARK, HENRY e JULNES, 2000).

Como as avaliações têm por finalidade subsidiar o entendimento sobre os programas e políticas, uma série de decisões devem ser tomadas com vistas a definir como serão estruturadas e conduzidas. Essas decisões afetam a extensão em que a avaliação poderá

fornecer informações para o aperfeiçoamento, correção, seleção ou entendimento sobre os programas e as políticas públicas (MARK, HENRY e JULNES, 2000).

Segundo ainda esses autores, o propósito final das avaliações é contribuir para a melhoria das condições sociais, sendo esse objetivo intermediado tanto pelas instituições democráticas, como pelos muitos interesses de seus defensores. Por essa razão, torna-se necessária a definição de propósitos mais delimitados para orientar e conduzir os trabalhos, com vistas ao alcance do fim maior (MARK, HENRY E JULNES, 2000).

Desta forma, desde que Scriven³ (1967) introduziu a clássica distinção entre os propósitos das avaliações – somativa e formativa – outros propósitos têm sido identificados pelos estudiosos, sendo o desenvolvimento de conhecimento o possível terceiro tipo, segundo Mark, Henry e Julnes (2000). Esse propósito tem como foco o desenvolvimento e/ou teste de proposições gerais sobre as causas dos problemas sociais, as soluções e o processo de construção das políticas, sem a preocupação direta com o aperfeiçoamento ou julgamento do programa estudado.

No entanto, não existe consenso sobre a tipologia de propósitos. Assim, esses autores sugerem uma quarta categoria: avaliações para o exame da conformidade e identificação de falhas, a qual foi adotada posteriormente por Patton (2008), conforme já mencionamos.

Desse modo, com vistas a demonstrar a relação entre os propósitos das avaliações sugeridos por Weiss (1998), Patton (2008) e Mark, Henry e Julnes (2000), apresenta-se, no QUADRO 1, uma visão comparativa para melhor contextualizar a interação entre eles.

Propósitos das Avaliações de Programas		
Weiss	Patton	Mark, Henry e Julnes
Corrigir o curso de um programa ou política com vistas a obter melhorias nos processos de execução.	Avaliações formativas - orientadas para o aperfeiçoamento, melhoria da qualidade, aprendizado organizacional e aperfeiçoamento contínuo.	Aperfeiçoamento de programas e organizações.
Documentar a história do programa de forma a contribuir para a sua compreensão e influenciar decisões futuras.		
Influenciar o comportamento dos executores por intermédio da ênfase nos objetivos do programa.		

(continua)

³ SCRIVEN, M. S. The Methodology of Evaluation. In: TYLER, R. W., GAGNE, R. M., SCRIVEN, M. S. (Ed). Perspectives of Curriculum Evaluation. Skokie, IL: AERA Monograph Series on Curriculum Evaluation, n. 1, p. 39-83, Rand MacNally, 1967.

(conclusão)

Propósitos das Avaliações de Programas		
Weiss	Patton	Mark, Henry e Julnes
Verificar o alcance dos objetivos com vistas à continuação, expansão ou institucionalização de um programa-piloto; assim como corte, finalização ou abandono de programa em execução.	Avaliações somativas - orientadas para julgamento de mérito e valor.	Julgamento do mérito e valor
Escolher a melhor alternativa, dentre várias opções, para implementar ações/medidas.		
Testar novas ideias sobre determinado programa	Avaliações desenvolvimentistas - orientadas para mudanças emergentes, inovações e transformações de processos.	Desenvolvimento de conhecimento
Compreender a intervenção social	Avaliações para produção de conhecimento - orientadas para identificar, através dos achados, os caminhos para a efetividade.	
Promover a <i>accountability</i>	<i>Accountability</i> - orientadas para explicar e justificar quão bem os recursos são gerenciados e se os resultados desejados são obtidos de forma eficiente.	Conformidade e Identificação de falhas
Contribuir como <i>feedback</i> para os executores disponibilizando informações sobre o andamento de determinado programa.	Monitoramento - orientadas para o acompanhamento da implementação com vistas a identificar tempestivamente os problemas (indicadores de desempenho e sistemas de monitoramento).	

QUADRO 1 - DEMONSTRATIVO DA RELAÇÃO ENTRE OS PROPÓSITOS DAS AVALIAÇÕES SUGERIDOS POR WEISS (1998), PATTON (2008) E MARK ET AL. (2000)

Para que as avaliações cumpram com os propósitos apresentados e os objetivos estabelecidos possam ser alcançados, pressupõe-se a utilização dos resultados. No entanto, o maior desafio para a avaliação, segundo Chen, “é produzir informações úteis para que o conhecimento e a tecnologia gerados possam resolver os problemas sociais e melhorar a qualidade de nossas vidas.” (CHEN *apud* ALKIN, 2004, p. 134, tradução nossa).

Conforme apontado por Alkin (2004), os estudos relacionados à utilização foram impulsionados pelas teorias orientadas para a decisão (*decision-oriented theories*). A partir dessas, as teorias da utilização foram construídas, cuja preocupação é conceber avaliações que não apenas contribuam para subsidiar o processo de tomada de decisão, como também assegurem que os seus resultados tenham impacto direto nas decisões e nas mudanças promovidas no âmbito das organizações e dos programas.

Para Stufflebeam (1994), um dos principais líderes do movimento pela padronização dos trabalhos de avaliação, para se examinar os propósitos das avaliações, deve-se observar os padrões geralmente aceitos para a realização dessa atividade. Nos Estados Unidos, esses

padrões foram estabelecidos e revisados pelo *Joint Committee on Standards for Educational Evaluation*, em 1994, conforme a seguir apresentado.

2.2 OS PADRÕES DE AVALIAÇÃO DE PROGRAMAS DO JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION

O Comitê Conjunto de Padrões para a Avaliação Educacional (*Joint Committee on Standards for Educational Evaluation*), responsável pela publicação dos Padrões de Avaliação de Programas, foi criado em 1975, nos Estados Unidos, a partir da junção das principais associações de profissionais preocupados com a qualidade da avaliação. Esse Comitê é credenciado pelo Instituto Nacional Americano de Padrões (*American National Standards Institute - ANSI*), sendo os padrões por ele aprovados considerados Padrões Nacionais Americanos. A sede do Comitê está localizada no Centro para Avaliação e Análise da Universidade de Iowa nos EUA (*JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION*, 2010).

Segundo Sanders (1994), os Padrões de Avaliação de Programas revisados e aprovados pelo Comitê em 1994 são o resultado de um projeto pioneiro iniciado em 1975, cujo objetivo era desenvolver padrões que ajudassem a assegurar a utilidade, viabilidade, ética e qualidade das avaliações de programas, projetos e materiais educacionais. No entanto, esses padrões devem ser utilizados, revisados e aperfeiçoados em todas as áreas, como parte de um esforço contínuo para se avançar na prática da avaliação.

Os padrões objetivam orientar as avaliações com vistas a assegurar a sua efetividade e buscam refletir as boas práticas da época em que foram aprovados.

Segundo Sanders (1994) “Um padrão é um princípio mutuamente aceito por pessoas engajadas em uma prática profissional, que, se observado, contribuirá para aumentar a qualidade e a fidedignidade daquela prática profissional, como por exemplo, a avaliação.” (SANDERS, 1994, p. 2, tradução nossa).

Com base na definição de Sanders (1994), tem-se que um padrão é um princípio geralmente aceito para uma determinada prática profissional, cujo objetivo é assegurar a qualidade e a credibilidade da atividade desenvolvida.

No Brasil, não existem padrões ou princípios estabelecidos para o exercício da atividade avaliativa. No entanto, considerando que os padrões estabelecidos pelo *Joint Committee* foram instituídos a partir do exercício dessa atividade, seus fundamentos são inerentes à prática da avaliação. Desta forma, mesmo não institucionalizados formalmente, os

princípios a serem observados quando da realização desse tipo de trabalho são tacitamente aceitos.

Os padrões de avaliação de programas instituídos pelo Comitê Conjunto foram organizados com base em quatro importantes atributos da avaliação: utilidade, viabilidade, propriedade e acuracidade. Para o Comitê, a observância desses atributos favorece a credibilidade e qualidade das avaliações (SANDERS, 1994).

Para cada um desses atributos foi definido um conjunto de padrões com vistas a assegurar o seu cumprimento; no entanto, alguns padrões são relevantes para mais de um atributo, mas, para efeito de organização dos mesmos, foram posicionados no grupo que reflete sua principal função.

2.2.1 Padrões de Utilidade

Os padrões de utilidade direcionam as avaliações, portanto devem ser informativos, tempestivos e influentes. Os avaliadores devem definir claramente aqueles atores que estarão envolvidos com a avaliação, como também, devem se assegurar das necessidades de informações desses atores, planejando as avaliações para atender essas necessidades e relatar as informações relevantes de forma clara e tempestiva.

Os padrões incluídos nesse atributo são: identificação dos *stakeholders*; credibilidade do avaliador; informações sobre a seleção e o escopo; identificação dos valores (critérios); relatório objetivo, tempestivo e disseminado; bem como avaliação do impacto. Em termos gerais, os padrões de utilidade definem se a avaliação atende, de forma prática, a necessidade de informações dos atores de um determinado público. O QUADRO 2 apresenta os padrões de utilidade e a sua descrição.

Padrões de Utilidade	Descrição
U1 - Identificação dos <i>stakeholders</i>	Identificar todos os atores envolvidos ou interessados que serão afetados pela avaliação, para que suas necessidades possam ser consideradas.
U2 - Credibilidade do avaliador	A avaliação deve ser conduzida por profissionais habilitados e competentes para exercício dessa atividade.
U3 - Definição do escopo	Deverão ser coletadas e selecionadas informações pertinentes para elaboração das questões que afetam o programa de forma que atendam às necessidades e aos interesses dos envolvidos (<i>stakeholders</i>).
U4 - Identificação dos valores	As perspectivas, procedimentos e lógicas utilizadas para interpretar os achados deverão ser cuidadosamente descritos, considerando que serão as bases para os julgamentos valorativos.
U5 - Clareza do relatório	Os relatórios de avaliação devem, objetivamente, descrever o que foi avaliado, incluindo o seu contexto, propósitos, procedimentos e achados, sendo essencial que forneçam informações para sua fácil compreensão.

(continua)

(conclusão)

Padrões de Utilidade	Descrição
U6 - Tempestividade e disseminação	Os achados significativos identificados no decorrer da avaliação e o relatório devem ser disseminados tempestivamente aos interessados na sua utilização.
U7 - Impacto da avaliação	As avaliações devem ser planejadas, conduzidas e reportadas de forma que encorajem o acompanhamento e participação dos stakeholders, para aumentar a probabilidade de utilização.

QUADRO 2 - PADRÕES DE UTILIDADE

FONTE: JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION, 1994, tradução nossa

2.2.2 Padrões de Viabilidade

Os padrões de viabilidade buscam assegurar que as avaliações serão realistas, prudentes, cordiais e econômicas. Desta forma, as avaliações devem ser concebidas considerando o ambiente e contexto natural do objeto avaliado. Para a execução dos trabalhos, não devem ser utilizados mais recursos, pessoas, materiais ou tempo do que o necessário para responder às questões estudadas.

Os três padrões desse atributo, apresentados no QUADRO 3, são: procedimentos práticos, viabilidade política e custo-efetividade.

Padrões de Viabilidade	Descrição
F1 - Procedimentos Práticos	Os procedimentos da avaliação devem ser práticos, mantendo o mínimo de interrupções enquanto as informações necessárias são obtidas.
F2 - Viabilidade Política	A avaliação deve ser planejada e conduzida antecipando-se as diferentes posições dos vários grupos interessados, de forma a se obter a cooperação de todos e, assim, possivelmente, evitar que qualquer desses grupos restrinja as atividades da avaliação, como também de modo a prevenir a utilização inadequada dos seus resultados.
F3 - Custo-Efetividade	A avaliação deve ser eficiente e as informações disponibilizadas devem ter suficiente valor, assim os recursos aplicados podem ser justificados.

QUADRO 3 - PADRÕES DE VIABILIDADE

FONTE: JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION, 1994, tradução nossa

2.2.3 Padrões de Propriedade (adequabilidade)

Os padrões de propriedade buscam assegurar que a avaliação será conduzida com observância aos princípios legais e éticos, respeitando o bem-estar das pessoas envolvidas na avaliação, assim como daquelas que podem ser afetadas pelos seus resultados.

Esses padrões pretendem facilitar a proteção aos direitos individuais que podem vir a ser infringidos por uma avaliação, de forma a proteger os envolvidos contra ilegalidades, condutas inescrupulosas e antiéticas, bem como ações ineptas por parte de todos aqueles que conduzem a avaliação.

Esses padrões requerem que os indivíduos que conduzem as avaliações aprendam sobre e obedeçam às leis relativas a questões como privacidade, liberdade de informações e proteção dos assuntos ligados às pessoas.

Os oito padrões desse atributo são: orientação para prestação dos serviços; acordo formal; assuntos relacionados a direitos individuais; relações humanas; análise completa e imparcial; apresentação dos achados; conflito de interesses e responsabilidade fiscal. A descrição desses padrões consta do QUADRO 4.

Padrões de Propriedade	Descrição
P1 - Orientação para Prestação dos Serviços	As avaliações devem ser concebidas para auxiliar as organizações a conduzir e efetivamente atender às necessidades do público alvo.
P2 - Acordo Formal	As obrigações das partes formais da avaliação devem ser documentadas, para que haja garantia de observância e cumprimento das condições acordadas.
P3 - Questões Relativas a Direitos Individuais	As avaliações devem ser concebidas e conduzidas respeitando e protegendo os assuntos relativos aos direitos individuais e ao bem-estar.
P4 - Relações Humanas	Os avaliadores devem respeitar a dignidade e valores em suas relações com outras pessoas associadas com a avaliação, de forma que os participantes não sejam ameaçados ou ofendidos.
P5 - Avaliação (julgamento) completa e imparcial	A avaliação deve ser completa e imparcial no exame e registro das forças e fraquezas do programa avaliado, assim pode-se avançar nos pontos fortes e encaminhar recomendações às áreas com problemas.
P6 - Apresentação dos achados	As partes formais de uma avaliação devem assegurar que a totalidade dos achados com as suas pertinentes limitações estarão acessíveis para as pessoas afetadas pela avaliação e quaisquer outras com expressos direitos legais de receber os resultados.
P7 - Conflitos de Interesses	Os conflitos de interesse devem ser tratados com clareza e honestidade de forma a não comprometer o processo e os resultados da avaliação.
P8 - Responsabilidade Fiscal	A alocação de recursos e as despesas realizadas pelo avaliador devem estar refletidas nos procedimentos de prestação de contas, caso contrário, deve-se ser prudente e eticamente responsável de forma que as despesas sejam comprovadas e registradas.

QUADRO 4 - PADRÕES DE PROPRIEDADE

FONTE: JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION, 1994, tradução nossa.

2.2.4 Padrões de Acuracidade

Os padrões de acuracidade buscam assegurar que as informações disponibilizadas pelas avaliações, sobre as características que determinam o valor ou mérito do programa, são tecnicamente adequadas e confiáveis.

Os avaliadores devem considerar tantas quantas forem as características práticas identificáveis do programa e devem coletar os dados daquelas características importantes, julgadas necessárias para analisar o mérito ou sucesso do programa. Além disso, as informações devem ser tecnicamente adequadas e a lógica da análise de mérito deve estar relacionada aos dados coletados.

Os doze padrões de acuracidade, descritos no QUADRO 45, são: documentação do programa; contextualização da análise; descrição dos procedimentos e propósitos; fontes de informações defensáveis; informações válidas; informações confiáveis; informações sistemáticas; análise quantitativa da informação; análise qualitativa da informação; conclusões justificadas; relatório imparcial e meta-avaliação. Esses padrões objetivam assegurar que a avaliação irá revelar e apresentar informações acuradas sobre o mérito ou sucesso do programa.

Padrões de Acuracidade	Descrição
A1 - Documentação do Programa	O programa avaliado deve ser descrito e documentado com acuracidade, de forma a que o programa possa ser claramente identificado.
A2 - Análise do Contexto	O contexto em que o programa está inserido deve ser examinado com bastante detalhe, com vistas a identificar a probabilidade de sua influência sobre o programa avaliado.
A3 - Descrição dos Propósitos e Procedimentos	Os propósitos e procedimentos da avaliação devem ser monitorados e descritos em detalhes, de forma que possam ser identificados e avaliados.
A4 - Fontes de Informações Defensáveis	As fontes de informações utilizadas na avaliação devem ser descritas em detalhes, de forma a possibilitar a avaliação da informação.
A5 - Informações Válidas	Os procedimentos para obtenção das informações devem ser escolhidos ou desenvolvidos, bem como aplicados, de forma que assegurem que a interpretação dos dados seja válida para os fins da utilização pretendida.
A6 - Informações Confiáveis	Os procedimentos de obtenção das informações devem ser escolhidos ou desenvolvidos, bem como aplicados, de forma a assegurar que as informações obtidas por meio deles sejam suficientemente consistentes para a sua utilização.
A7 - Informações Sistematizadas	As informações coletadas, processadas e relatadas na avaliação devem ser sistematicamente revisadas e qualquer erro encontrado deve ser corrigido.

(continua)

(conclusão)

Padrões de Utilidade	Descrição
A8 - Análise Quantitativa da Informação	As informações quantitativas de uma avaliação devem ser adequada e sistematicamente analisadas de forma a responder efetivamente às questões da avaliação.
A9 - Análise Qualitativa da Informação	As informações qualitativas em uma avaliação devem ser adequada e sistematicamente analisadas de forma que as questões da avaliação possam ser efetivamente respondidas.
A10 - Conclusões Justificadas	As conclusões apresentadas em uma avaliação devem ser explicitamente justificadas para que os <i>stakeholders</i> possam avaliá-las.
A11 - Relatório Imparcial	Os procedimentos para elaboração do relatório devem salvaguardar contra a ocorrência de distorções decorrentes de sentimentos pessoais e outros de qualquer natureza, de forma que o relatório reflita fielmente a avaliação dos achados.
A12 - Meta-avaliação	A própria avaliação deve ser formativa e somativamente avaliada contra esses e outros padrões pertinentes, de forma a garantir que foi adequadamente conduzida e orientada e, em complemento, os <i>stakeholders</i> podem pertinentemente examinar suas forças e fraquezas

QUADRO 5 - PADRÕES DE ACURACIDADE

FONTE: JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION, 1994, tradução nossa

Sanders (1994) ressalta que esses padrões são princípios orientadores e não regras, devendo ser utilizados como meio para permitir a troca de informações sobre a qualidade e credibilidade da avaliação entre os avaliadores, seus clientes e outros *stakeholders*.

No entanto, os padrões auxiliam, ainda, na discussão de uma série de questões. Dentre elas, a análise dos propósitos da avaliação, os quais, segundo o Padrão A3 (QUADRO 5), são entendidos nos termos dos objetivos da avaliação e na intenção de utilização de seus resultados (SANDERS, 1994).

Para Weiss, “a utilidade de alguma forma assegura a racionalidade para as avaliações, mesmo que o uso não seja direto e imediato”. (WEISS, 1998, p.15, tradução nossa). Mas, o que é utilização?

2.3 DEFINIÇÃO DE UTILIZAÇÃO

Para Levinton e Hughes (1981), a utilização se refere ao uso dos resultados das avaliações no âmbito de programas e políticas, observando-se dois critérios. Primeiro, deve existir uma relação entre os achados e as questões examinadas nos programas e políticas, cujos resultados possam ser traduzidos nas implicações para essas questões; e, segundo, para ser considerado utilizado deve haver evidências de que, na ausência das informações

produzidas pelas avaliações, todos aqueles engajados nas atividades do programa ou da política teriam agido de forma diferente.

Os critérios servem para distinguir entre utilização e impacto. Este pode ser entendido como as modificações no programa ou na política para as quais os achados da avaliação contribuíram.

O modelo apresentado por esses autores contempla de forma bastante abrangente a modalidade de utilização orientada pelos achados das avaliações. No entanto, as evidências demonstram que quanto mais se aprofundou no estudo do contexto, cultura, estrutura organizacional e política, surgiu uma nova modalidade de utilização, a qual se tem denominado uso no processo, que pode ser definida como fonte de influência da avaliação e se refere aos efeitos produzidos, em termos de mudanças de atitudes e ações, durante o curso da avaliação, estimulado pela participação dos usuários no processo de execução dos trabalhos (MARK e HENRY, 2004).

Desta forma, alguns autores têm destacado que o conceito de utilização tem sido expandido para incluir novas e intangíveis formas de influenciar indivíduos, programas e comunidades (Kirkhart, 2000; Henry e Mark, 2003; Mark e Henry, 2004; Lawrenz *et al.*, 2007). Os proponentes desse movimento consideram que a teoria das avaliações influentes contempla de forma mais adequada as consequências da avaliação, como aquelas que não estão limitadas pelo tempo e contexto (LAWRENZ *et al.*, 2007, p. 276).

A utilização, portanto, pode ser expressa nos meios e mecanismos que favorecem o alcance dos resultados. Leviton (2003) destaca a abordagem apresentada por Kirkhart (2000) “que deixa explícito que o uso da avaliação opera dentro de múltiplos domínios e serve a uma variedade de funções ao mesmo tempo. Sua estrutura é especialmente útil para elucidar como os executores dos programas podem usar a avaliação, e por intermédio de quais canais ocorre a sua utilização.” (Leviton, 2003, p. 525, tradução nossa). Outro exemplo é a abordagem de Mark e Henry (2004), “que utiliza valiosos conceitos das ciências sociais para identificar os processos de utilização nos níveis individual, interpessoal e coletivo.” (LEVITON, 2003, p. 525, tradução nossa).

No entanto, conforme as considerações de Patton (2008), “não existe uma definição absoluta e ideal para utilização, porque isso depende em parte dos valores e objetivos daqueles que utilizarão as informações produzidas pela avaliação.” (PATTON, 2008, p. 102, tradução nossa).

Lawrenz *et al.* (2009) na apresentação realizada na Conferência da Associação Americana de Avaliação, em 12 de novembro de 2009, sob o título “*Relationships between*

Involvement and Use in the Context of Multi-site Evaluation”⁴ sintetizou no slide abaixo transcrito (QUADRO 6) a definição dos termos: utilização da avaliação, influência sobre a avaliação e influência da avaliação.

Termo	Definição
Utilização da Avaliação	A aplicação objetivada dos processos, achados ou conhecimento para produzir um efeito.
Influência sobre a Avaliação	A capacidade de um indivíduo produzir efeitos sobre uma avaliação por meios diretos ou indiretos.
Influência da Avaliação (Kirkhart, 2000)	A capacidade ou força da avaliação para produzir efeitos sobre os outros por meios indiretos ou intangíveis.

QUADRO 6 - DEFINIÇÕES DOS TERMOS UTILIZAÇÃO E INFLUÊNCIA SEGUNDO LAWRENZ *ET AL.*, 2009, TRADUÇÃO NOSSA

A utilização, portanto, é o meio por intermédio do qual o alcance dos resultados esperados é favorecido, sendo que a utilização pode ser mensurada a partir das dimensões a seguir apresentadas.

2.4 DIMENSÕES DA UTILIZAÇÃO

A literatura apresenta uma distinção clássica para as categorias de utilização das informações produzidas pelas avaliações segundo os propósitos que as originaram, que é amplamente utilizada para fundamentar as pesquisas nesse tema e a partir das quais outras abordagens têm sido sugeridas. Segundo Shulha e Cousins (1997) a utilização, ao invés de ser um conceito unitário, pode ser considerada um fenômeno multidimensional, melhor descrito pela interação de três dimensões segundo os propósitos a que se destinam: instrumental, conceitual e simbólica (ou política).

A dimensão instrumental, definida por Rich e segundo Leviton e Hughes (1981), diz respeito aos casos onde os resultados da avaliação são utilizados diretamente para tomada de decisão ou para solução de problemas relacionados aos achados reportados nas avaliações.

Esse tipo de utilização atende a uma das premissas básicas da avaliação, influenciar, direta e imediatamente, as decisões com base nos achados identificados. No entanto, segundo Weiss, essa dimensão pura não é comum. “Muitos estudos não são utilizados dessa forma, por que os decisores são influenciados por uma série de outros fatores, como, por exemplo, as expectativas dos participantes do programa e seus gestores, o apoio dos financiadores, as demandas do poder, os custos da mudança, a capacidade de realização, dentre outras, que acabam por frustrar a sua utilização imediata” (WEISS *et al.*, 2005, p. 13, tradução nossa).

⁴ Relações entre o envolvimento e a utilização no contexto da avaliação de múltiplas localidades.

No entanto, ainda segundo Weiss, o uso instrumental é comum sob três condições:

- a) se as implicações dos achados não forem controversas, nem provocarem rupturas na organização, tampouco produzirem conflito de interesses;
- b) se as mudanças que serão implementadas estiverem dentro do repertório existente do programa e forem relativamente em pequena escala; e,
- c) se o ambiente do programa for relativamente estável, sem grandes mudanças na liderança, orçamento, tipo de clientes ou público atendido.

Além dessas, há uma quarta condição que favorece a utilização instrumental: quando o programa está em crise ou paralisado e ninguém sabe o que fazer (WEISS, 1998a).

A dimensão conceitual caracteriza-se quando os resultados das avaliações são utilizados para influenciar a opinião de pessoas sobre determinado programa ou política, para introduzir determinado conhecimento, ideia ou conceito, que são úteis para dar sentido à determinada proposta, sem que exista qualquer ação ou decisão em decorrência dos achados produzidos pelas avaliações. No entanto, esses achados constituem importantes informações, em abstrato, para utilização futura (Patton, 2008). Esse tipo de influência é caracterizada como “*enlightenment*” (WEISS).

Alguns autores têm tratado a dimensão conceitual como educacional, aprendizado organizacional (Feldman, 1989 *apud* Weiss *et al.*, 2005) ou processo cognitivo (Cousins e Leithwood, 1986 *apud* Weiss *et al.*, 2005). Teóricos ao redor do mundo têm evidenciado, em pesquisas e avaliações, que a dimensão conceitual tem sido a mais efetiva forma de influenciar a política (Anderson e Biddle, 1991; Baklien, 1983; Ballart, 1998; Bienayme, 1984; Bulmer, 1987; Husen e Kogan, 1984; Radaelli, 1995 *apud* Weiss *et al.*, 2005). Segundo Weiss, quando os achados das avaliações são introduzidos na arena decisória, direta ou indiretamente, algumas vezes e a longo prazo, passam a se constituir em um novo conhecimento comum (WEISS *et al.*, 2005, p.14).

A dimensão simbólica (ou política) é utilizada para legitimar decisões ou posicionamentos preexistentes. A avaliação pode ser utilizada para justificar decisões que seriam tomadas de qualquer forma. Nesta dimensão, a avaliação serve como instrumento de persuasão (WEISS *et al.*, 2005; WEISS, 1998a).

Uma quarta dimensão tem sido apresentada na literatura, denominada uso no processo (*process use*), atribuída a Patton. Em decorrência da aplicação da teoria focada na utilização (*utilization-focused evaluation*), enfatiza o efeito da participação dos usuários interessados nos resultados durante todo o processo de avaliação. Esse modelo tem como

premissa realizar avaliações úteis com base nas intenções de uso para os interessados na utilização. (*intended use by intended users*).

No entanto, Weiss a considera muito mais uma fonte de uso do que propriamente uma nova dimensão. Aliás, para Weiss, as dimensões clássicas parecem capturar muito da experiência empírica e prática. Pois, o que se observa é que o uso no processo explicita a ideia que tem estado implícita por muito tempo, mas não é paralela às outras três dimensões. O uso conceitual serve para a compreensão, o uso político subsidia e justifica a adoção ou não de determinada ação e o uso no processo nos fala como a avaliação influencia as ações no decorrer da sua realização (WEISS *et al.*, 2005).

Para Mark e Henry, no entanto,

[...] essas categorias de utilização são caracterizadas por diferentes atributos qualitativos. As utilizações instrumental e conceitual são definidas em termos do tipo de mudança a ser produzida em decorrência da avaliação. Para o uso instrumental, as mudanças ocorrem nas ações, enquanto que, no uso conceitual, as mudanças ocorrem no âmbito pessoal afetivo-cognitivo (percepções, sentimentos). Na utilização simbólica, ao contrário, o resultado da avaliação não é diretamente mensurado pelos seus efeitos, mas pode ser percebido nos propósitos de atores ou organizações, como os políticos, fazendo uso dos achados das avaliações para justificar uma posição preexistente, ou quando uma determinada agência é vista como utilizando a avaliação para parecer racional e responsável. Por outro lado, o uso no processo é definido como fonte de influência da avaliação; refere-se aos efeitos produzidos em decorrência da participação durante o processo de avaliação e não daqueles decorrentes dos seus achados. Pode-se presumir que a participação no processo de avaliação poderá contribuir para mudar tanto atitudes como ações, e desta forma, o uso no processo poderia se sobrepor às utilizações conceitual e instrumental. (MARK e HENRY, 2004, p. 36, tradução nossa).

Para acomodar a complexidade desse construto, Kirkhart (2000) propôs três dimensões para caracterizar a influência: fonte (processo e resultados), intenções (intencionais e não intencionais) e tempo (durante, ao término da avaliação ou no futuro); cujo modelo, segundo Weiss é uma forma de integrar o uso no processo (fonte) com outros aspectos da utilização (intenção e tempo).

Recentemente, outros pesquisadores têm defendido que o objetivo último da avaliação deveria ser a melhoria das condições sociais (*social betterment*), especialmente para os beneficiários dos programas sociais, mediante a determinação do bem comum, da seleção e adaptação do curso de ação (Christie, 2007; Henry & Mark, 2003; Mark & Henry, 2004; Mark, Henry e Julnes, 1999 *apud* Lawrenz *et al.*, 2007). Para tanto, Mark & Henry (2004) propuseram um modelo teórico de avaliação influente estruturado a partir do modelo lógico proposto por Cousins (2003, *apud* Mark e Henry, 2004), considerando as entradas (*inputs*), que se referem tanto ao contexto da avaliação como ao da decisão/política; as atividades inerentes à avaliação; os produtos (*outputs*) gerados pela avaliação e os mecanismos mediante

os quais se proporciona a integração desses produtos aos resultados intermediários e de longo prazo, com vistas ao alcance do objetivo maior da avaliação, que é a promoção da melhoria social.

Na avaliação de Carol Weiss, Mark & Henry

“[...] apresentaram como premissas um conjunto de categorias teóricas - mediadores e alternativas - através dos quais a avaliação exerce influência. Concebido a partir da ciência social, eles desenvolveram a “teoria da mudança” para referenciar as consequências da avaliação nos níveis individual, interpessoal e coletivo. Por exemplo, no nível das ações organizacionais (eles preferem o termo ações coletivas), pressupõem que a avaliação pode exercer influência por intermédio dos processos de definição de agenda, aprendizado orientado para a política, mudança na política ou difusão. Esse trabalho recombina algumas das categorias existentes da utilização da avaliação e tenta estabelecer o processo mediante o qual a influência é exercida, tanto conceitualmente distinta, como melhor integrada. O esquema proposto por esses autores contempla alternativas para se compreender a influência, como também sistematizam as pesquisas sobre o tema. (WEISS, 2005, p. 14, tradução nossa).

Ao conduzir uma avaliação no Programa de Educação Contra o Abuso de Drogas (*Drug Abuse Resistance Education - D.A.R.E.*), Weiss (2005) identificou os mesmos mecanismos descritos por Mark & Henry (2004), tendo ainda encontrado outro em que a avaliação afetou as ações organizacionais, o qual denominou uso impositivo, que não é um conceito novo, mas pode ocorrer em qualquer campo onde os altos níveis do governo estabelecem condicionalidades para que os níveis operacionais recebam recursos. O que o distingue dos demais tipos de utilização é que os *stakeholders* são obrigados a observar os resultados das avaliações. No caso do *D.A.R.E.*, eles poderiam perder os recursos se não concordassem em adotar o programa nos moldes em que cientificamente tem se provado efetivo.

Pode-se concluir que a tipologia de utilização amplamente aceita no campo desta pesquisa pode ser sintetizada nas dimensões: instrumental, conceitual e simbólica e o uso no processo (QUADRO 7).

Tipo	Utilizado para	Definição: a utilização do conhecimento...
Instrumental	Agir	...para tomar decisões
Conceitual ou Enlightenment	Compreender	...para melhor compreender um programa ou política
Político, Persuasivo ou Simbólico	Justificar	...para suportar decisões já tomadas ou para persuadir outros para uma opinião específica

QUADRO 7 - TIPOLOGIA DE UTILIZAÇÃO SEGUNDO LAWRENZ *et al.*, 2009, TRADUÇÃO NOSSA

No entanto, em que pese se tenha antecipado muito da discussão teórica sobre a utilização, com vistas à melhor contextualização do tema, mostrou-se relevante inserir, nesta revisão de literatura, as discussões que antecederam e propiciaram as novas abordagens sobre

a utilização dos resultados das avaliações. A seguir, iniciamos com o debate sobre esse assunto entre dois eminentes teóricos Carol Weiss e Michael Patton.

2.5 RESGATANDO A DISCUSSÃO SOBRE A UTILIZAÇÃO DA AVALIAÇÃO: O DEBATE WEISS-PATTON (1988)

Em 1987, na reunião nacional da Associação Americana de Avaliação, Carol Weiss iniciou um debate memorável com Michael Patton – posteriormente reproduzido em artigos do *American Journal of Evaluation* – que se tornou um marco na discussão e no desenvolvimento de pesquisas sobre a utilização dos resultados das avaliações de programa, ante a relevância do tema para o aperfeiçoamento da teoria e prática da avaliação.

Weiss, em seu clássico artigo intitulado “*Is Anybody There? Does Anybody Care?*” (1988a), pondera, com base nas evidências dos estudos avaliativos, que os resultados das avaliações pouco influenciavam, em decorrência dos seus achados, as decisões relativas à implementação de melhorias no âmbito dos programas, quer seja no nível gerencial, ou na esfera política. Segundo essa autora, “usualmente, as coisas parecerem continuar muito mais como elas seriam se as avaliações não tivessem sido feitas.” (WEISS, 1988a, p. 5, tradução nossa). Essa constatação está fundamentada em cinco razões:

- i) pouca atenção é dada à natureza política do processo de decisão no âmbito dos programas;
- ii) as avaliações não são abrangentes e convincentes o bastante para fornecer a “resposta correta”;
- iii) muitas das decisões relativas aos programas não obedecem a um processo racional de tomada de decisão;
- iv) mesmo quando existem decisões objetivas e de fácil implementação pelos respectivos responsáveis, as pessoas não sabem que tipo de informação é necessária; e, finalmente,
- v) existe uma constelação de interesses que “protege” o *status quo* do programa.

Patton, naquela oportunidade, discordando da abordagem de Weiss, ressaltou que “[...] para a pesquisa política, as conclusões de Weiss podem ter mérito, mas para a avaliação de programas tem-se tido muito sucesso e se deve e pode-se fazer melhor.[...]” (PATTON, 1988, p.10, tradução nossa).

Ao tratar das divergências, Patton reiterou que existiam entre ele e Weiss visões, paradigmas e percepções diferentes de como conduzir os trabalhos de avaliação, tanto quanto

aos seus propósitos, responsabilidades e o que realisticamente se deveria alcançar. Na opinião dele, Weiss considerou apenas quatro possibilidades para a utilização dos resultados da avaliação no processo político, quais sejam:

- i) identificar possíveis problemas no desempenho do programa;
- ii) estabelecer diretrizes para melhoria do programa ou da política;
- iii) fornecer subsídios para interpretação das atividades e dos resultados; e ainda,
- iv) mobilizar apoio para novas propostas.

Ele adiciona outras, como aquelas que diretamente influenciam as mudanças no âmbito dos programas, que afetam imediatamente as decisões, como por exemplo, a adoção de medidas corretivas e/ou aperfeiçoadoras durante o processo de execução das avaliações (uso no processo).

Para Patton, amparado nos princípios definidos pelo *Joint Committee on Standards for Educational Evaluation*, o mérito da avaliação é a sua utilização, “[...] as avaliações destinam-se a contribuir para o aperfeiçoamento dos programas, para melhorar a sua efetividade, fornecendo informações para que decisões sejam tomadas. Isso é uso. Isso é a responsabilidade do avaliador.” (PATTON, 1988, p. 14, tradução nossa).

Em resposta a Patton, Weiss no artigo intitulado “*If Program Decisions Hinged Only on Information: A Response to Patton*” (1988b) reafirmou que, quanto à utilização, o que o avaliador pode aspirar a alcançar é influenciar mudanças no programa. A avaliação deveria ser um instrumento para o contínuo aprendizado dos formuladores e gestores dos programas, assim como dos políticos.

Ademais, Weiss reforça a natureza política das avaliações, que havia sido refutada por Patton, enfatizando que preferências e interesses individuais, que permeiam o processo decisório mesmo nas organizações, precedem a importância que deveria ser dada aos resultados das avaliações, quando colocados em risco esses valores.

Cabe observar, no entanto, que na última edição do livro “*Utilization-Focused Evaluation*”, (2008), Patton, 20 anos depois desse debate, fazendo referência ao desafio colocado por Weiss, reconhece que existem limitações para o uso das informações produzidas pelas avaliações em relação ao que elas podem ou não oferecer; que os interesses dos tomadores de decisão podem estar além das evidências apresentadas; que a politização e a distorção dos achados, o mau uso das avaliações e a excessiva retórica da *accountability* são fatores que afetam os benefícios que deveriam ser por elas produzidos. “A avaliação, como qualquer ferramenta, pode ser utilizada para o bem e para o mal” (PATTON, 2008, p.xvii,

tradução nossa), por esta razão os avaliadores têm um papel adicional em fazer com que a retórica seja aderente à realidade, buscando assegurar o uso apropriado da avaliação.

2.6 A UTILIZAÇÃO DA AVALIAÇÃO: A REVISÃO TEÓRICA E PRÁTICA DE SHULHA E COUSINS (1997)

Shulha e Cousins (1997), dentre as iniciativas do Grupo Temático de Interesse na utilização da avaliação (*Evaluation Use TIG*) da Associação Americana de Avaliação, promoveram uma ampla revisão e síntese da literatura sobre esse tema publicada desde 1986. Observaram, naquela oportunidade, avanços no desenvolvimento das teorias, que incluíam o surgimento de considerações sobre o contexto como sendo crítico para o entendimento e explicação da utilização; a identificação do uso no processo como uma consequência significativa da atividade avaliativa; bem como a expansão do conceito de utilização, passando do indivíduo para o nível organizacional, e a diversificação do papel do avaliador, como facilitador, planejador e educador. Além disso, um novo foco para o desenvolvimento teórico foi identificado, que poderia limitar a extensão das pesquisas e se referia à compreensão da utilização inadequada das avaliações (má utilização).

Para esses autores, desde 1986, há um consenso de que o uso ou a utilização não é um conceito unitário, mas sim, um fenômeno multidimensional melhor caracterizado pela interação das dimensões instrumental, conceitual e simbólica, que pode ser mensurado pela análise de fatores como relevância, credibilidade, envolvimento do usuário, comunicação, processamento da informação, atendimento às necessidades de informação dos clientes, antecipação de mudanças nos programas, percepção da avaliação como instrumento de gerenciamento, qualidade da avaliação e características do contexto e da decisão ou política.

Para Shulha e Cousins, a pesquisa de Cousins e Leithwood (1986) – que será apresentada a seguir – e as observações de Greene (1988) redirecionaram as pesquisas no sentido de se construir teorias, já que promover a utilização não é simplesmente descrever diferentes tipos e uso e catalogar os fatores que contribuem para a sua efetivação. No entanto, a contribuição mais importante para o desenvolvimento teórico e que introduziu a importância do papel do contexto nas pesquisas, segundo esses autores, surgiu justamente do debate Weiss-Patton, apresentado anteriormente, que trouxe à discussão a questão da responsabilidade do avaliador em assegurar a utilização das avaliações,

para Weiss as informações produzidas pelas avaliações competem para dar credibilidade às decisões em um contexto bem caracterizado como não-racional e pluralístico. Desta forma, o muito que o avaliador pode e deve almejar é apresentar achados acurados e adequados às questões investigadas. Patton, por outro lado,

ênfatisa o papel do avaliador em promover e incentivar a utilizaç o, sendo responsabilidade dos avaliadores, tanto contribuir para que os tomadores de decis o identifiquem o que precisa ser avaliado, como determinar, de forma participativa, qual informaç o melhor atende a suas necessidades e, ainda, apresentar os achados que atendam ao tipo de informaç o demandada por aqueles que t m intenç o de utiliz -las. (SHULHA E COUSINS, 1997, p. 197, traduç o nossa).

Para Smith e Chircop *apud* Shulha e Cousins (1997), a import ncia do debate Weiss-Patton est  no confronto de diferentes abordagens em termos de percepç o dos programas e dos interesses dos atores envolvidos. Assim, sugerem que a teoria da utilizaç o deveria diferenciar entre os contextos em que os *stakeholders* discutem os objetivos do programa (contexto pol tico), daqueles em que os objetivos j  foram definidos e o objeto da investigaç o refere-se aos meios para alcanç -lo (contexto da tomada de decis o racional).

Para Shulha e Cousins (1997), em s ntese, o que se observou foi que as teorias e pesquisas sobre a utilizaç o das avaliaç es deram novas dimens es ao contexto, em investigaç es multifacetadas com componentes epistemol gicos, pol ticos e organizacionais, tendo-se reconhecido que todas as avaliaç es s o pol ticas e todos os achados merecem atenç o.

Segundo esses autores, alguns proeminentes te ricos, como Guba e Lincoln (1989) *apud* Shulha e Cousis (1997), defendiam uma abordagem construtivista da avaliaç o, em que a criaç o do conhecimento est  integrada ao contexto em que ele   produzido. Assim, como as avaliaç es criam conhecimento, o ato de avaliar   utilizar.

J  uma segunda abordagem enfatiza que a atividade pol tica est  intrinsecamente relacionada ao contexto, onde pessoas, ambiente e circunst ncias est o fortemente conectados, influenciando a utilizaç o. Nessa perspectiva, considera-se que o tomador de decis o n o age sozinho e enfrenta uma  rdua competiç o por informaç es relevantes com outros grupos de interesse (GREENE, 1990; KING, 1988, *apud* SHULHA e COUSINS, 1997; WEISS, 1988a e 1988b).

A terceira abordagem prioriza a relaç o entre estrutura e processos organizacionais com a utilizaç o, segundo a qual, a complexidade dos n veis burocr ticos dentro das organizaç es, as linhas de comunicaç o dentro e entre esses diversos n veis e os mecanismos dominantes para interpretaç o dos significados das informaç es produzidas pelas avaliaç es, contribuem para a potencial utilizaç o dos achados (JENLINK, 1994; MATHISON, 1994; OWEN e LAMBERT, 1995; PRESKILL, 1991, 1994a; PRESKILL e TORRES, 1996; TORRES, PRESKILL e PIOTNEK, 1996 *apud* SHULHA e COUSINS, 1997).

Já outros pesquisadores e teóricos, preocupados com a importância dos fatores contextuais, dedicaram-se mais diretamente a compreender a relação entre o avaliador e o contexto do programa. Segundo Shulha e Cousins, a contribuição de Patton segue essa linha, considerando que as avaliações devem servir às intenções de uso dos interessados na utilização.

Essa questão, inclusive, despertou a atenção de alguns autores interessados na perspectiva social e psicológica do aprendizado, reconhecendo o poder da construção social do significado das informações das avaliações.

Alguns estudos inovadores sobre a utilização têm procurado, ainda, compreender o impacto da avaliação considerando as dimensões cognitiva, afetiva e política no processo de utilização (GREENE, 1988; PRESKILL e TORRES, 1996; COUSINS e EARL, 1992; FETTERMAN, 1994 *apud* SHULHA e COUSINS, 1997).

2.7 A PESQUISA SOBRE UTILIZAÇÃO DAS AVALIAÇÕES DE LEVITON E HUGHES (1981)

Na seção anterior, examinamos a revisão da teoria, pesquisa e prática da utilização da avaliação, elaborada por Shulha e Cousins (1997), que resgatou a contribuição do debate Weiss-Patton e contextualizou os novos rumos que se tem tomado no estudo da utilização.

A partir desta seção, iniciamos a revisão dos trabalhos, em ordem cronológica, que abordam especificamente os fatores e mecanismos que influenciam a utilização das avaliações. Apresentamos, a princípio, o estudo elaborado por Leviton e Hughes (1981), cujo objetivo foi revisar as definições e as metodologias adotadas, àquela época, para o estudo da utilização das avaliações de programas e políticas, com vistas a identificar as variáveis que favorecem a utilização. O trabalho elaborado pelos autores foi estruturado em três partes: discussão crítica da definição de utilização; discussão metodológica; e, revisão das variáveis que favorecem a utilização.

Ao tratar da definição de utilização, os autores enfatizaram que a definição por eles apresentada se aplicaria apenas ao uso dos resultados das avaliações de programas e políticas, sendo distinta de impacto. Para tanto, a partir do estabelecimento de dois critérios: evidência de que os resultados podem ser traduzidos em recomendações sobre as questões investigadas (processamento da informação, segundo WEISS, 1978 e WEINBERG, 1979 *apud* LEVITON e HUGHES, 1981); e evidências de que, na ausência das informações produzidas pelas avaliações, os atores envolvidos na política ou nas atividades do programa teriam agido de forma diferente, levantam as hipóteses de que as avaliações podem ter pouca utilidade, porque

são de baixa qualidade ou suas implicações podem não ser exequíveis; ou ainda, sua utilidade pode não ser reconhecida ou as pessoas podem preferir não utilizá-las. Finalmente, as avaliações podem não ter impacto, porque isso depende das decisões de muitas pessoas no governo, que podem falhar em perceber a sua utilidade, ou escolher não utilizar as avaliações.

Com relação às questões metodológicas, os autores destacam que são muitos os problemas relacionados aos métodos adotados no estudo da utilização, em razão das estratégias de pesquisa aplicadas, que evidenciam deficiências na conceitualização de utilização, como também a baixa prioridade na documentação das evidências. Muitos estudos adotam a abordagem metodológica de estudo de caso ou entrevistas e *surveys*. Esses métodos, segundo os autores, enfrentam pelo menos quatro problemas: a dificuldade em documentar que a utilização ocorreu, considerando que as avaliações são frequentemente utilizadas informalmente, bem como ante o tempo decorrido entre a utilização e a realização do estudo; a dificuldade em demonstrar que a mudança, em qualquer nível, foi causada pelo menos em parte pela avaliação; a questão das bases para comparação; e, por último, a identificação da unidade de análise, ou seja, o que seria uma instância de utilização?

Para Leviton e Hughes, se queremos demonstrar que as avaliações são mais utilizadas, ou que a utilização das informações e resultados das avaliações aumentou, é necessário mensurar (quantificar) a utilização.

No que diz respeito às variáveis que favorecem a utilização, essas foram classificadas em cinco categorias vinculadas à relevância, as formas e canais de comunicação, a forma de processamento das informações (compreensão), a credibilidade dos avaliadores e da avaliação, como também a participação dos usuários, conforme apresentado no QUADRO 8.

Variáveis	Atributos
Relevância	A. Avaliação direcionada para as necessidades dos clientes B. Necessidades dos políticos C. Necessidades dos gestores dos programas D. Tempestividade
Comunicação	A. Comunicação direta dos usuários sobre as necessidades para avaliação B. Comunicação por intermédio da hierarquia da burocracia C. Disseminação
Processamento de Informações	A. Reconhecimento da relevância B. Apresentação das informações C. Processamento de informações de acordo com o estilo dos administradores
Credibilidade	A. Comparação com outras informações B. Preconceitos dos usuários C. Credibilidade do avaliador D. Qualidade
Participação e Apoio dos Usuários	A. Comprometimento com a avaliação B. Apoio ao programa e políticas

QUADRO 8 - CONJUNTO DE VARIÁVEIS E ATRIBUTOS RELACIONADOS À UTILIZAÇÃO
FONTE: LEVITON E HUGHES (1981)

Para cada um dos atributos que afetam a utilização das avaliações, Leviton e Hughes identificaram um conjunto de fatores para mensurar as variáveis que estão relacionadas no QUADRO 9: relevância, comunicação, processamento de informações, credibilidade, participação e apoio dos usuários.

O trabalho de Leviton e Hughes destaca-se pela iniciativa em sistematizar os fatores que afetam a utilização em categorias (variáveis), a partir das quais se investigou a sua importância para a utilização. Os resultados demonstram que as variáveis são representativas para indicar aqueles fatores que favorecem ou dificultam a utilização.

No entanto, dadas as fragilidades nos instrumentos de coleta de dados utilizados nas pesquisas que foram estudadas pelos autores, principalmente quanto à documentação das evidências para suportar as conclusões, Leviton e Hughes destacam a necessidade de se melhorar os métodos para se mensurar a utilização.

Não obstante essa constatação, para cada um dos atributos que afetam a utilização das avaliações, Levinton e Hughes relacionaram um conjunto de fatores. Aqueles que mais contribuíram para a utilização das avaliações foram marcados com um sinal positivo (+), por outro lado, aqueles assinalados com o sinal negativo (-) não favoreceram a utilização. Quando assinalados com os sinais mais ou menos (+/-), significa que, associados a outros fatores, podem propiciar ou não a utilização (QUADRO 9).

Variáveis	Atributos
Relevância	<p>(+) A. Avaliação direcionada para as necessidades dos clientes</p> <p>B. Necessidades dos políticos</p> <p>(+) informações sobre alocação de recursos</p> <p>(+) informações sobre implementação</p> <p>(+) informações gerais sobre a efetividade</p> <p>C. Necessidades dos gestores dos programas</p> <p>(+) informações sobre implementação</p> <p>(+) informações sobre a efetividade dos elementos do programa</p> <p>(-) informações sobre a efetividade geral do programa tem baixa prioridade</p> <p>D. Tempestividade</p> <p>(+/-) provável interação com o tipo de uso</p> <p>(+/-) interação com o estágio do ciclo da política</p> <p>(+) considerações prévias sobre as evidências</p>
Comunicação	<p>(+) A. Comunicação direta dos usuários sobre as necessidades para avaliação</p> <p>(-) B. Comunicação por intermédio da hierarquia da burocracia</p> <p>C. Disseminação</p> <p>(+) disseminação direta</p> <p>(-) informações excluídas dos sumários</p> <p>(-) distorções ou deficiências de interpretação nos sumários - possível má utilização</p> <p>(-) informações negativas suprimidas - possível má utilização</p> <p>(+) disseminação em rede</p> <p>(-) disseminação entre agências</p> <p>(+) revisão dos pares ("invisible colleges of specialists")</p>

(continua)

(conclusão)

Variáveis	Atributos
Processamento de Informações	<p>A. Reconhecimento da relevância (+) avaliação responde a questões específicas (-) objetivos não são claramente especificados (-) as avaliações locais são relevantes para o nível Federal (+/-) informações não antecipadas</p> <p>B. Apresentação das informações (+) apresentação clara (-) jargões aumentam a dificuldade</p> <p>C. Processamento de informações de acordo com o estilo dos administradores (-) atenção somente para as informações relevantes (-) baixa preocupação com qualidade - possível má utilização (+) comunicação verbal (-) deficiências quanto à utilização de métodos - possível má utilização (+/-) preferência por informações qualitativas - interação com o conteúdo da avaliação (-) diferenças entre os estilos administrativo e acadêmico</p>
Credibilidade	<p>A. Comparação com outras informações (-) contradições entre pesquisas e outras informações (-) contradições entre intuições e expectativas dos usuários (+) informações corroborativas (+) utilização adicional de pesquisas</p> <p>B. Preconceitos dos usuários (+) opiniões fortes sobre pesquisa (+) confiança na pesquisa sobre a intuição (+) usuários com formação médica (-) usuários com formação em direito (uso persuasivo) (+) usuários estão temporariamente no governo</p> <p>C. Credibilidade do avaliador (-) provável suspeição de cooptação (-) provável utilização fora da agência</p> <p>D. Qualidade (-) baixa percepção de qualidade (-) baixa qualidade - possível má utilização (+/-) interação com outros tipos de utilização - convencer outros (+) metodologia não questionada facilmente</p>
Participação e Apoio dos Usuários	<p>A. Comprometimento com a avaliação (+) elevado interesse e comprometimento dos usuários (+) apoio a informações produzidas pelas avaliações (+/-) provável interação entre o comprometimento dos usuários e as mudanças na política (+/-) provável interação entre o comprometimento dos usuários e a rotatividade destes</p> <p>B. Apoio ao programa e políticas (+/-) interação entre os rumos dos achados (+) avaliações menos visíveis (+) menor quantidade de deliberadores (+/-) provável interação entre os resultados da avaliação e sua tempestividade</p>

QUADRO 9 - CONJUNTO DE VARIÁVEIS E DETALHAMENTO DOS ATRIBUTOS RELACIONADOS À UTILIZAÇÃO

FONTE: LEVITON E HUGHES (1981)

2.7.1 Análise dos fatores de Leviton & Hughes

Segundo o estudo realizado por Leviton e Hughes (1981), foi possível aferir a relevância dos fatores por eles sistematizados, que influenciam a utilização, conforme demonstraremos a seguir.

a) Relevância

Segundo Leviton e Hughes (1981, p. 534) muitos autores têm destacado que as avaliações são utilizadas com mais frequência quando têm alta relevância para os problemas que comprometem os resultados das políticas ou dos programas. Para Lynn (1978) *apud* Leviton e Hughes (1981) a dificuldade de se identificar a relevância está relacionada a duas questões: atender às necessidades dos clientes (políticos e gestores) e a tempestividade da informação.

Muitos estudos relacionam diretamente a utilização com o atendimento à necessidade de informações dos clientes (alta relevância). Assim como a baixa relevância, a baixa utilização. Resnick *et al.* *apud* Leviton e Hughes comprovaram que as avaliações são utilizadas com mais frequência quando são relevantes para atender as necessidades de um público em particular. Além disso, o estudo de Resnick *et al.* demonstrou divergências entre a utilização por políticos e gestores. Outros estudos demonstraram que tanto os políticos como os gestores consideram úteis as informações práticas sobre a implementação. No entanto, enquanto o estudo de Weiss e Bucuvala *apud* Leviton e Hughes concluiu que, de um modo geral, para os políticos todas as considerações sobre a efetividade dos programas são úteis, o trabalho de Nielsen *apud* Leviton e Hughes considerou que os gestores preferem informações sobre a efetividade de diferentes elementos do programa. Leviton e Hughes argumentam, ainda, que existem estudos que evidenciam que trabalhos sobre a efetividade em geral dos programas não são prioritários para muitos gestores, como também para aqueles que prestam os serviços.

Quanto à segunda questão, a tempestividade, muitos autores têm veementemente enfatizado a importância de se submeter os resultados das avaliações em tempo para a decisão (BANTA e BAUMAN, 1976; FLACONE e JAEGER, 1976; GUBA, 1975 *apud* LEVITON e HUGHES, 1981). Mas ressaltam que dois estudos concluíram que a tempestividade não é consistentemente importante para a utilização dos resultados. Young e Comptois, 1979 *apud*

Leviton e Hughes sugeriram que a tempestividade é importante para o uso instrumental a curto prazo, mas menos importante para o uso conceitual.

b) Comunicação

Leviton e Hughes apontam que dois estudos demonstraram a importância da boa comunicação entre os avaliadores e os potenciais usuários dos resultados da avaliação, sendo inclusive um indicador do sucesso do trabalho. No entanto, a comunicação dentro das burocracias tende a ser obstaculizada. Downs, (1967) *apud* Leviton e Hughes (1981), depois de um amplo estudo, concluiu que o “homem-médio” dentro da hierarquia burocrática transmite seletivamente a informação entre os escalões superiores e inferiores. Assim, a informação pode ser eliminada ou distorcida. Desta forma a comunicação por meio da hierarquia da burocracia pode afetar negativamente a utilização, porque o avaliador pode falhar em obter as informações completas e isentas sobre as necessidades dos clientes.

c) Processamento de informações

Um dos critérios estabelecidos pelos autores para mensurar a utilização da avaliação é que a relevância seja reconhecida e que os achados possam ser traduzidos nas implicações para o que é relevante.

Assim, a forma como os resultados das avaliações são apresentados para os usuários afeta a sua compreensão e, desta forma, a extensão em que serão utilizados. Alguns autores sugerem que os relatórios escritos são mais utilizados, como também que as recomendações explícitas podem favorecer a utilização. De um modo geral, a comunicação clara facilita a disseminação.

Outro ponto destacado pelos autores diz respeito ao estilo dos gestores no processamento das informações (COX, 1997 *apud* LEVITON e HUGHES, 1981), que pode gerar as seguintes implicações: somente os achados relevantes para as necessidades dos gestores serão recebidos com atenção; possivelmente, não estarão sujeitos a uma avaliação de qualidade; e, a comunicação verbal favorecerá a utilização, porque os gestores preferem ser informados continuamente.

Os administradores preferem informações qualitativas a quantitativas. Além disso, a diferença de estilo entre os avaliadores e os gestores pode afetar a utilização.

d) Credibilidade

Os gestores utilizam uma série de indicadores para valorar as informações disponibilizadas pelas avaliações, que são os indicadores de credibilidade. Esses indicadores estão relacionados no QUADRO 9. As Avaliações são utilizadas para corroborar outras informações, de forma a obter ganhos de informação em termos agregados como mostrado por Patton *et al.*, 1977 e Rich *apud* Leviton e Hughes.

A credibilidade do avaliador também parece ser importante. Mas o mais importante parece ser a suspeita de que os avaliadores tenham sido cooptados ou tenham suprimido informações. Segundo os autores, Guba (1975) adverte para o fato de que muitos avaliadores desejariam produzir as informações que os clientes querem.

Outro ponto que afeta a credibilidade diz respeito à alta qualidade metodológica, que apesar de alguns autores defenderem esse critério, a qualidade não tem sido consistentemente encontrada afetando a utilização. A qualidade metodológica parece influenciar mais a utilização quando a informação é essencial para o convencimento. No estudo de Weiss e Bucuvalas (1977), a qualidade foi relacionada a dois tipos de utilização: para mobilizar apoio para um posicionamento e para mudar o entendimento sobre determinada questão.

e) Participação e apoio dos usuários

O apoio é esperado no contexto político. No entanto, o apoio para as informações produzidas pelas avaliações concorre com outros tipos de apoio. A relação entre avaliações e apoio é, em grande parte, o contexto político em que a avaliação é conduzida. Tanto na pesquisa de Patton *et al.*, 1977, como na de Caplan *et al.*, 1975, *apud* Leviton & Hughes (1981) há consenso de que esse contexto é essencial para se compreender a utilização. Para outros autores, o apoio contínuo de um indivíduo-chave é fundamental para ter os achados da avaliação utilizados, assim como o compromisso do tomador de decisão determina se a avaliação será utilizada.

Patton, (1978) *apud* Leviton & Hughes (1981), concluiu que a utilização é favorecida quando a avaliação é direcionada para um indivíduo em particular que é receptivo ao trabalho. No entanto, as vantagens dessa abordagem são limitadas por dois fatores da burocracia: as mudanças repentinas na política e a alta rotatividade a que estão sujeitos os altos escalões no governo.

A próxima pesquisa destacada, nesta revisão, foi realizada por Cousins e Leithwood (1986), e teve por objetivo analisar os fatores que influenciam a utilização das avaliações. Nesse estudo, os autores calcularam um índice das relações entre a utilização (variável

dependente) e um conjunto de fatores (variáveis independentes) relacionados a duas categorias (contextos da implementação da avaliação e da decisão/política), conforme veremos a seguir.

2.8 A PESQUISA SOBRE UTILIZAÇÃO DAS AVALIAÇÕES DE COUSINS E LEITHWOOD (1986)

O trabalho de Cousins e Leithwood (1986) teve por objeto a revisão da pesquisa empírica sobre a utilização dos resultados de 65 avaliações, abrangendo o período de 1971 a 1985, com o objetivo de analisar os fatores que influenciavam a utilização das avaliações. O estudo buscou responder às seguintes questões:

- i) Quais as características metodológicas dos estudos empíricos para investigar a utilização da avaliação e quais são seus determinantes?
- ii) Como as variáveis dependentes têm sido operacionalizadas nesses estudos?
- iii) Quais as orientações que têm sido adotadas pelas variáveis independentes?
- iv) Que fatores têm sido apresentados que afetam a natureza da utilização dos resultados das avaliações?

Para o desenvolvimento da pesquisa, os autores adotaram as seguintes definições:

- a) resultados das avaliações: qualquer informação associada com o impacto (*outcomes*) da avaliação. Por exemplo: dados, interpretações, recomendações;
- b) objeto da avaliação: currículo, programa, projeto, escola, curso, pessoal, ou estudantes;
- c) tomadores de decisão: afiliados com um ou alguns dos serviços. Por exemplo: educação, saúde, saúde mental ou serviço social;
- d) variável dependente: utilização, entendida como, i) subsídio para decisões discretas e ii) aprendizado dos tomadores de decisão. Além dessas duas possibilidades, utilizou-se também o processamento psicológico como modalidade de utilização, sem que, necessariamente, se tenha influenciado decisões, correspondido a ações ou mudanças de entendimento;
- e) variáveis independentes: fatores que influenciam a utilização dos resultados das avaliações pelos tomadores de decisão. Os autores utilizaram duas categorias para agrupar os fatores que influenciam a utilização, aqueles relacionados às

características dos contextos: i) da implementação das avaliações; e, ii) da decisão ou da política.

Para realização do trabalho, os autores utilizaram as pesquisas retrospectiva, longitudinal e de simulação.

Com relação às conclusões da pesquisa, a utilização foi relacionada à tomada de decisão, ao aprendizado e ao processamento das informações. Alguns estudos não foram muito claros com relação ao tipo de utilização, evidenciando uma utilização potencial.

Com relação à utilização entendida como subsídio para decisões discretas, foram relacionados três tipos diferentes de decisões adotadas no âmbito dos programas avaliados: i) decisões sobre alocação de recursos; ii) decisões sobre a natureza ou forma de execução; iii) decisões sobre o gerenciamento.

No que diz respeito às variáveis independentes, essas foram classificadas em duas categorias, as relacionadas à implementação da avaliação e as referentes às características da decisão ou política, conforme apresentado no QUADRO 10.

Variáveis Independentes – Fatores	
Implementação da avaliação	Decisão/política
Qualidade da avaliação	Informações necessárias
Credibilidade	Características da decisão
Relevância	Clima político
Comunicação (relatório)	Informações concorrentes
Achados	Características pessoais
Tempestividade	Compromisso e/ou receptividade da avaliação

QUADRO 10 - VARIÁVEIS INDEPENDENTES SEGUNDO COUSINS E LEITHWOOD, 1986

Por intermédio da FIGURA 1, pode-se visualizar a estrutura conceitual da utilização da avaliação proposta por Cousins e Leithwood, que se baseia na combinação analítica das variáveis dependentes e independentes apresentada no estudo realizado. O construto utilização demonstra que os resultados da avaliação são, psicologicamente, processados pelos tomadores de decisão antes da formação do conceito ou da tomada de decisão. As doze variáveis independentes ou fatores foram agrupados em duas categorias hipotéticas. A divisão em duas categorias indica que são institutos dinâmicos interagindo entre si para produzir a utilização.

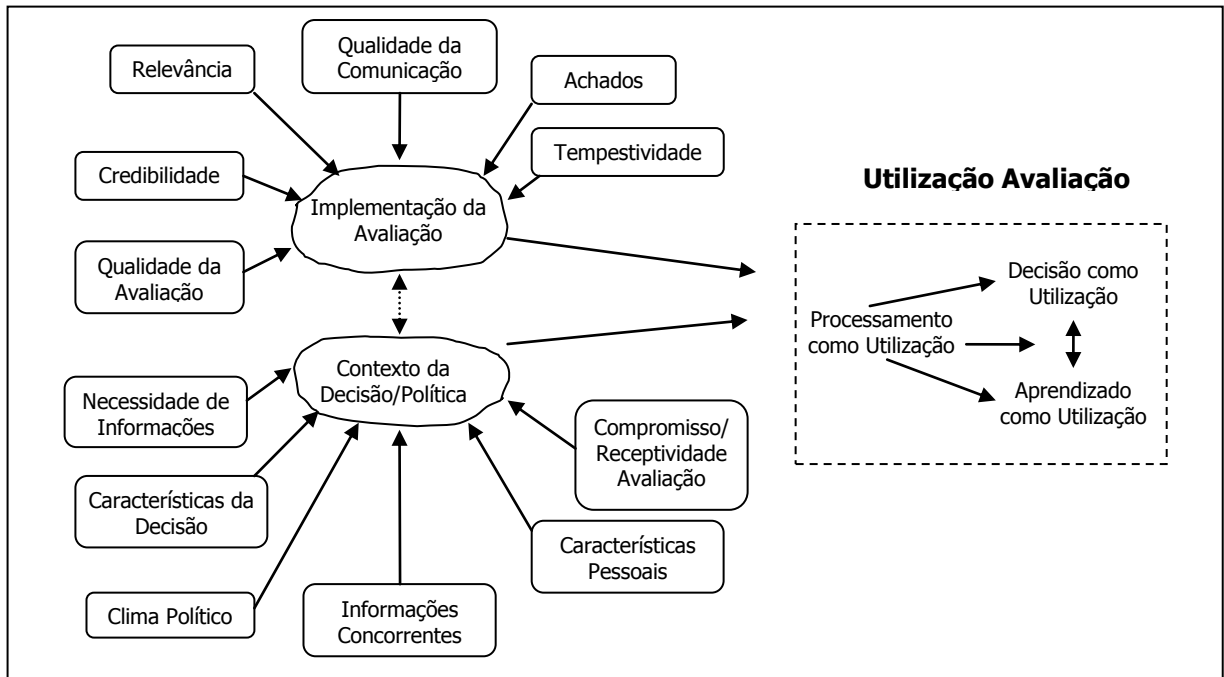


FIGURA 1 - ESTRUTURA DA UTILIZAÇÃO DA AVALIAÇÃO DE COUSINS E LEITHWOOD (1986, P 348, tradução nossa).

O estudo de Cousins e Leithwood revelou que a influência relativa dos 12 fatores varia em função do tipo de uso. De um modo geral, a utilização das avaliações parece ser mais evidente quando:

- i) a abordagem da avaliação foi adequada, assim como a sofisticação metodológica e a sua representatividade;
- ii) as decisões adotadas pelos usuários foram significativas e consideradas apropriadas em comparação com os dados coletados;
- iii) os achados foram consistentes com os valores e expectativas dos usuários;
- iv) os usuários foram envolvidos no processo de avaliação e estavam comprometidos com os benefícios da avaliação;
- v) os usuários consideravam relevantes os dados do relatório de avaliação para os seus problemas;
- vi) um percentual mínimo de informações de outras fontes conflitavam com os resultados da avaliação (COUSINS E LEITHWOOD, 1986, p. 360, tradução nossa).

O modelo meta-analítico (FIGURA 1) proposto por Cousins e Leithwood (1986) introduziu, na pesquisa sobre a utilização, os fatores relacionados ao contexto político, passando a contemplar os valores, interesses, preferências e conflitos que influenciam a utilização das avaliações.

No entanto, as conclusões da pesquisa demonstraram que, em que pese a relevância dos fatores relacionados ao contexto político, a qualidade, sofisticação e representatividade da

metodologia utilizada estavam entre os fatores mais importantes para influenciar a utilização, conforme será apresentado a seguir.

2.8.1 Análise dos fatores de Cousins e Leithwood (1986)

Conforme mencionamos, o trabalho de Cousins e Leithwood (1986) acrescentou à discussão sobre utilização uma segunda categoria de fatores, que são aqueles relacionados ao contexto político/decisório, que, segundo demonstrado na literatura, também são importantes para favorecer a utilização das informações e resultados disponibilizados pelas avaliações.

No entanto, pode-se depreender que os fatores relacionados à categoria da implementação da avaliação já haviam sido identificados no trabalho realizado por Leviton e Hughes (1981). Comparativamente, podemos visualizar os fatores conforme apresentado no QUADRO 11.

Categorias	Fatores	
	Leviton e Hughes	Cousins e Leithwood
Implementação da Avaliação	Relevância	Relevância
	Comunicação	Comunicação
	Processamento de Informações	Qualidade da avaliação
	Credibilidade	Achados
Credibilidade		
Tempestividade		
Decisão/Política	Participação/Apoio dos Usuários	Compromisso/Receptividade da avaliação
		Necessidade de Informações
		Clima político
		Informações concorrentes
		Características pessoais
		Características da decisão

QUADRO 11 - COMPARAÇÃO FATORES QUE INFLUENCIAM A UTILIZAÇÃO. APRESENTADOS POR LEVITON & HUGHES (1981) E COUSINS E LEITHWOOD (1986).

Para melhor compreensão da pesquisa empreendida por Cousins e Leithwood (1986), apresenta-se, a seguir, a análise dos fatores realizada pelos autores.

No que diz respeito à categoria implementação das avaliações, os autores ressaltaram que a identificação dos fatores a ela relacionados foram evidentes em estudos que exploraram as seguintes questões:

- i) A metodologia aplicada na avaliação parece confiável?
- ii) A avaliação é relevante para atender às necessidades do tomador de decisão declaradas ou não declaradas?
- iii) Os resultados da avaliação foram apresentados de forma inteligível e são consistentes com as expectativas dos tomadores de decisão?
- iv) Os resultados foram disponibilizados em tempo para a tomada de decisão?

Os fatores relativos à implementação das avaliações foram definidos pelos autores da seguinte forma:

- a) Qualidade da avaliação: características do processo de avaliação, incluindo a sofisticação metodológica, rigor e tipo do modelo de avaliação.
- b) Credibilidade do avaliador e do processo de avaliação, definido em termos de objetividade, acreditação, adequação dos critérios da avaliação, dentre outros.
- c) Relevância da avaliação para as necessidades de informação do tomador de decisão em termos do propósito da avaliação e da vinculação organizacional do avaliador.
- d) Qualidade da comunicação: clareza na apresentação dos resultados da avaliação para os respectivos públicos, em termos de estilo, confiança do avaliador nos resultados e extensão da disseminação.
- e) Achados: positivos e negativos, consistentes com as expectativas dos públicos da avaliação, importantes para a tomada de decisão, dentre outras.
- f) Tempestividade na disseminação dos resultados da avaliação para o tomador de decisão.

Os fatores relacionados ao contexto da decisão/política, segundo Cousins e Leithwood (1986), não dizem respeito apenas às características organizacionais, mas também são as necessidades de informações de todos os públicos das avaliações e envolvem questões como:

- i) Quais são as características do público e em que extensão os usuários da avaliação estão envolvidos com o objetivo da avaliação?
- ii) Quais são as características do objetivo da avaliação?
- iii) Quais são os recursos disponíveis?
- iv) Existem outras fontes de informação, como propostas de políticas ou experiências prévias, que concorrem com as informações que emergem da avaliação?
- v) Quem são os principais tomadores de decisão e quais são as atitudes deles em relação à avaliação ou pesquisa de um modo geral?
- vi) Como as decisões são tomadas e a estrutura é flexível?

Estão associados a essas questões os seis fatores a seguir descritos, conforme apresentados pelos autores:

- a) Necessidade de informações dos públicos da avaliação, incluindo o tipo de informação, número de públicos da avaliação com diferentes necessidades de informação, tempo e percepção da necessidade da avaliação.
- b) Características da decisão: área de impacto, tipo de decisão, programa novo e a significância da decisão ou do problema da avaliação, dentre outras.
- c) Clima político: orientação política dos responsáveis pela avaliação, grau de dependência dos tomadores de decisão com os financiadores externos, rivalidades inter e intra-organizacional, batalhas orçamentárias, conflitos de poder, dentre outras.
- d) Concorrência de informações de outras fontes, além da avaliação (observações do pessoal, staff, pares, etc), sobre o problema, concorrendo com as informações da avaliação.
- e) Características pessoais definidas em termos dos papéis dos tomadores de decisão na organização, estilo de processamento das informações, experiência na organização, características sociais, dentre outras.
- f) Comprometimento e/ou receptividade com a avaliação: atitudes dos tomadores de decisão durante a avaliação, resistência, abertura, dentre outras.

2.8.2 Os resultados da análise dos fatores de Cousins e Leithwood (1986)

Com base no estudo elaborado pelos autores, foi possível mensurar a relevância dos fatores quanto a sua capacidade de influenciar a utilização conforme a seguir demonstrado.

a) Qualidade da avaliação

Com relação à qualidade da avaliação, em 40% dos estudos foi observada uma relação entre qualidade e uso. A utilização na maioria das evidências coletadas é oposta ao uso potencial. A qualidade foi tipicamente definida como sofisticação metodológica, tipo de abordagem do problema, ou representatividade das atividades da avaliação.

Para muitos autores, a sofisticação metodológica foi positivamente relacionada à utilização em termos de suporte para decisões e o desenvolvimento conceitual dos usuários.

As avaliações com foco no processo ou na implementação do programa foram em geral consideradas mais úteis que aquelas que focaram estritamente nos resultados (*outcomes*).

Barnette e Thompson, 1979, *apud* Cousins e Leithwood (1986) observaram que a representatividade (definida como frequência das observações) favoreceu a utilização dos achados para decisões instrumentais.

b) Credibilidade

Mais de um quarto dos estudos examinados pelos autores demonstram uma relação entre utilização e credibilidade do avaliador ou do processo de avaliação (geralmente definido em termos de adequação dos critérios da avaliação).

Segundo os autores, quando as avaliações são percebidas pelos tomadores de decisão como altamente válidas ou quando enfatizam atividades importantes, a utilização e o potencial para utilização parecem ser grandes. No entanto, quando a coleta de dados é percebida como inapropriada por esses decisores a sua associação com o uso é reduzida.

c) Relevância

Muitos estudos demonstraram que as avaliações que refletiam o contexto em que seriam utilizadas – atendiam às preferências dos tomadores de decisão, apresentavam consenso sobre o problema da avaliação ou demonstravam estar inseridas nas operações dos programas e no processo de decisão – foram associadas com altos níveis de utilização. E, avaliações que ignoraram essas questões foram associadas com baixos níveis relativos de utilização.

Na mensuração da relevância, a vinculação do avaliador (interno ou externo) com a organização apresentou resultados conflitantes.

No entanto, as relações realizadas pelos autores sugeriram que as avaliações internas são mais utilizáveis que as externas.

Avaliações internas são aquelas realizadas por profissionais do próprio programa ou da própria organização avaliada e as avaliações externas são aquelas realizadas por pessoas de fora. No entanto, uma avaliação pode ser considerada tanto interna como externa, como, por exemplo, quando uma “instituição de saúde manda uma equipe da sede para avaliar o programa em seis diferentes estabelecimentos.” Neste caso, para aqueles que se encontram nos estabelecimentos é externa, mas para os administradores do órgão central é interna (WORTHEN *et al.*, 2004).

d) Comunicação

Um terço das avaliações estudadas pelos autores tratou da relação entre utilização e estilo de comunicação, atividades de comunicação em curso e extensão da disseminação.

As evidências sobre estilo de comunicação e utilização foram inconclusivas. Em contraste, as atividades de comunicação em curso e a proximidade do avaliador do tomador de decisão favoreceram a utilização, assim como a confiança do avaliador nos resultados e a intensidade da comunicação.

e) Achados

Muitos estudos demonstraram que, quando os achados das avaliações são coerentes com as expectativas dos tomadores de decisão, sua aceitação e utilização são favorecidas; achados incoerentes tendem a ser ignorados e outras informações são utilizadas.

Os achados das avaliações foram reportados como sendo muito úteis para algumas propostas, como: legislação, desenvolvimento organizacional e outros tipos de decisões em nível local ou federal, especialmente quando os achados eram práticos e conclusivos e quando identificados cursos de ação alternativos para os tomadores de decisão.

f) Tempestividade

Menos de 15% dos estudos examinados demonstraram relação entre utilização e tempestividade na comunicação dos resultados das avaliações. No entanto, a disponibilização dos resultados das avaliações foi positivamente relacionada com o índice de utilização. A tempestividade é relevante para o uso instrumental de curto prazo.

g) Necessidade de informações

Quando os tomadores de decisão percebem o grau de importância da informação, o uso dos resultados é favorecido e sua tendência é concordar com as recomendações das avaliações.

David, 1978 *apud* Cousins e Leithwood (1986) observou que os programas com alta estabilidade estão associados com o não uso dos resultados das avaliações para o aperfeiçoamento dos programas. Por outro lado, Brickell (1976) encontrou que a grande necessidade teve um efeito adverso na preparação do relatório da avaliação.

O tipo de informação requerida pelos tomadores de decisão também influencia a utilização, quando essa não é coerente com as informações da avaliação, a relevância do trabalho é reduzida.

A diversidade das necessidades dos públicos foi considerada negativamente relacionada com a utilização dos resultados para os propósitos da decisão.

h) Características da decisão

Muitos dos estudos examinaram a utilização da avaliação para decisões em diferentes áreas. A utilização foi maior em áreas relativas a pontos fortes e fracos dos programas, à faculdade de aperfeiçoamento, a facilidades no gerenciamento e na gestão do programa. Também se mostrou muito útil para utilização nos estágios iniciais do processo de decisão, quando as alternativas de solução estão sendo construídas.

i) Clima político

A influência política na utilização da avaliação ocorreu tanto no nível organizacional como extraorganizacional. Somente dez estudos observaram uma relação entre os fatores políticos e a utilização da avaliação.

Dentro da organização, as visões existentes do *staff*, os arranjos organizacionais e recompensas modelam a avaliação e influenciam a extensão em que será utilizada. A diminuição do uso ocorre se os resultados da avaliação forem percebidos como uma ameaça pelo *staff*, se eles sentirem que “estão com as mãos amarradas”, se os membros-chave do *staff* deixarem a organização ou se existirem debates e disputas orçamentárias internos. Estão associadas também a diminuição do uso, as rivalidades entre agências e as pressões sobre os avaliadores exercidas pelos diretores e executores dos programas.

King e Pechman, 1984 *apud* Cousins e Leithwood (1986) apontaram que as exigências do nível federal, para decisões sobre alocação de recursos, favorecem a utilização da avaliação para promover a *accountability*.

j) Informações concorrentes

O termo conhecimento do trabalho, atribuído a Kennedy, 1983, 1984 *apud* Cousins e Leithwood (1986) que é derivado da experiência pessoal, crenças, valores, interesses e objetivos, foi apresentado como sendo um concorrente das informações da avaliação. Muitos estudos mostraram que o conhecimento decorrente da observação pessoal reduziu a importância da avaliação para o tomador de decisão.

Outros autores demonstraram que o processamento de muitos conjuntos de informação reduz a dependência do tomador de decisão da avaliação.

k) Características pessoais

Mais de um terço da amostra estudada pelos autores demonstrou relação e não-relação entre a utilização da avaliação e as características pessoais do tomador de decisão, incluindo o seu papel na organização, treinamento, experiência e as características da liderança.

Os achados sobre as relações entre o papel da pessoa dentro da hierarquia organizacional e a utilização da avaliação foram conflituosos. Assim como a relação entre utilização e experiência foi ambígua. No entanto, as características da liderança foram consideradas positivamente relacionadas ao uso da avaliação em muitos estudos.

l) Comprometimento e/ou receptividade da avaliação

Aproximadamente um quarto dos estudos examinados pelos autores tratou do efeito na utilização, do comprometimento dos tomadores de decisão com a avaliação e/ou suas atitudes durante os trabalhos. Muitos desses estudos demonstraram a conexão direta com a utilização.

O alto envolvimento contribui para elevados níveis de utilização em termos de impacto e na formulação de políticas, assim como favorece o aprendizado organizacional sobre o programa, a programação de decisões a serem tomadas e aumenta, em certa medida, outros fatores, como a comunicação, relevância, processamento de informações e a credibilidade.

Segundo as conclusões do estudo de Cousins e Leithwood (1986), os fatores que mais influenciam a utilização são a qualidade da avaliação e as características da decisão, seguidas dos achados, do comprometimento e/ou receptividade da avaliação e da relevância. Os tipos de utilização mais frequentes, associados aos fatores foram para decisão e processamento.

Recentemente Johnson *et al.* (2009) realizaram uma revisão da literatura empírica sobre a utilização da avaliação, compreendendo o período de 1986 a 2005, na qual utilizaram a estrutura do trabalho de Cousins e Leithwood (1986) para categorizar os estudos sobre a utilização da avaliação. Como conclusão do trabalho, os autores identificaram uma nova categoria e um novo fator relacionado ao contexto da implementação da avaliação, que influenciam a utilização: o envolvimento dos *stakeholders* (categoria) e a competência do avaliador (fator). Os achados demonstram a importância do envolvimento dos *stakeholders* para facilitar a utilização da avaliação e sugere que o engajamento, a interação e a

comunicação entre os clientes da avaliação e os avaliadores são críticos para valorar a utilização das avaliações.

2.9 A PESQUISA SOBRE A UTILIZAÇÃO DA AVALIAÇÃO - UMA REVISÃO DA LITERATURA EMPÍRICA DE 1986 A 2005 DE JOHNSON *ET AL.*(2009)

Para seleção dos trabalhos a serem examinados na pesquisa realizada por Johnson *et al* (2009), diferentemente daquelas realizadas por Cousins e Leithwood (1986) e Shulha e Cousins (1997), foram identificados, no período examinado, mais de 600 estudos empíricos que tratavam da utilização da avaliação, dentre artigos, relatórios e capítulos de livros, e 48 dissertações, sendo selecionados, ao final, apenas 41, que atenderam ao padrão mínimo de qualidade definido pelos autores, que permitiria a comparação com o trabalho anterior de Cousins e Leithwood (1986). Dos estudos selecionados, 38 tratavam da utilização dos achados e três do uso no processo.

Cabe observar que, para Johnson *et al.* (2009), a utilização da avaliação foi definida como a aplicação dos processos, produtos ou achados da avaliação para produzir um efeito (JOHNSON *et al.*, 2009, p. 378, tradução nossa).

Os autores utilizaram como critérios para avaliar a qualidade dos trabalhos uma adaptação dos critérios propostos por Guarino, Santibañez, Daley e Brewer (2004) e Guba e Lincoln (1989), que consideraram alguns aspectos como: definição clara do problema, adequabilidade do desenho da pesquisa, consistência entre as evidências apresentadas e as conclusões, bem como em que extensão a margem de erro pode ter comprometido os resultados. A equipe de pesquisadores também avaliou o tamanho da amostra, a seleção e mensuração das variáveis e a interpretação estatística dos estudos quantitativos, assim como a adequação metodológica, transparência, qualidade descritiva e padrões utilizados nos estudos qualitativos. Os estudos empíricos foram classificados em cinco níveis: pobre, adequado-baixo, adequado-consistente, adequado-alto e excelente.

Quanto aos resultados da pesquisa, os autores constataram que a competência do avaliador é o novo fator identificado na categoria das variáveis independentes relacionadas à implementação da avaliação. Dos 41 estudos revisados, 6 indicaram que as características dos avaliadores sugeriam que esses profissionais desempenham um importante papel na condução da avaliação para que elas sejam utilizadas. Muito embora no trabalho de Cousins e Leithwood (1986), tenha sido considerado o fator credibilidade, no sentido da sua reputação técnica (escolhas de métodos e critérios), essa definição não abrange as características

relativas à competência e liderança, que se mostraram importantes na influência que exercem sobre a utilização. O que os autores ressaltam é que a nova característica da competência do avaliador tem uma relação maior com a pessoa do avaliador (quem ele é).

A outra contribuição do trabalho de Johnson *et al.* foi a identificação de uma nova categoria de variáveis independentes: o envolvimento dos *stakeholders*. A inclusão dessa nova categoria ao trabalho original de Cousins e Leithwood (1986) é um reflexo, segundo os autores, da ampliação da utilização das abordagens de avaliações participativas desde 1985.

Nessa nova categoria, os autores identificaram nove características, sendo que oito delas se parecem com aquelas identificadas por Cousins e Leithwood, mas sendo complementadas com o envolvimento (QUADRO 12). Muito embora o trabalho anterior incluísse o envolvimento dos *stakeholders* no fator comprometimento e/ou receptividade com a avaliação pertencente à categoria da decisão/política, no trabalho de Johnson *et al.* (2009), 23 dos 41 estudos relacionavam o envolvimento com a utilização, sugerindo tratar-se de uma nova categoria.

Categorias	Fatores	
	Cousins e Leithwood (1986)	Johnson et al. (2009)
Implementação avaliação	Relevância	Envolvimento com relevância
	Comunicação	Envolvimento com qualidade da comunicação
	Qualidade da avaliação	
	Achados	Envolvimento com achados
	Credibilidade	Envolvimento com credibilidade
	Tempestividade	Competência do avaliador
Decisão/Política	Compromisso/Receptividade da avaliação	Envolvimento com compromisso ou receptividade da avaliação
	Necessidade de informações	Envolvimento com necessidade de informações
	Clima político	
	Informações concorrentes	
	Características pessoais	Envolvimento com características pessoais
	Características da decisão	Envolvimento com características da decisão
Envolvimento stakeholders		

QUADRO 12 - RELAÇÃO ENTRE OS FATORES QUE FAVORECEM A UTILIZAÇÃO - IDENTIFICADOS POR COUSINS E LEITHWOOD (1986) E JOHNSON ET AL. (2009)

Em muitos estudos revisados, o envolvimento foi identificado como facilitador do processo de avaliação, que, em decorrência, contribuiu para aumentar a implementação da avaliação. Em outros, o envolvimento contribuiu para decisões ou definição de políticas que promoveram uma maior capacidade para utilização das informações disponibilizadas pelas avaliações. Pode-se inferir que o envolvimento dos *stakeholders* é um mecanismo que facilita o processo ou a definição da avaliação, facilitando a comunicação, o que leva a uma maior utilização. Os achados evidenciam, ainda, que a comunicação entre os avaliadores e seus

clientes é um fator crítico para a importância da utilização das avaliações (Johnson et al., 2009)

Característica	Descrição da característica
Envolvimento com compromisso ou receptividade da avaliação	O envolvimento dos stakeholders com a avaliação cria compromisso ou receptividade da avaliação.
Envolvimento com qualidade da comunicação	O envolvimento dos stakeholders promove a melhoria da comunicação.
Envolvimento direto dos stakeholders	A relação direta entre o envolvimento e a utilização da avaliação.
Envolvimento com credibilidade	O envolvimento dos stakeholders aumenta a credibilidade do processo e/ou do avaliador.
Envolvimento com achados	O envolvimento dos stakeholders com a avaliação em conhecer e entender os achados da avaliação.
Envolvimento com relevância	A participação dos stakeholders para integrar ao desenho da avaliação problemas importantes da organização.
Envolvimento com características pessoais	O envolvimento na avaliação de stakeholders de diferentes níveis da organização.
Envolvimento com características da decisão	O envolvimento de stakeholders de diferentes níveis, dependendo das características da decisão que precisa ser tomada.
Envolvimento com necessidade de informações	O envolvimento dos stakeholders facilita a introdução de suas necessidades de informações.

QUADRO 13 - DESCRIÇÃO DAS CARACTERÍSTICAS (FATORES) DO ENVOLVIMENTO DOS STAKEHOLDERS IDENTIFICADAS POR JOHNSON ET AL. (2009)

Os autores argumentaram que foi impossível responder à questão sobre que características (fatores) mais contribuíram para a utilização de uma forma direta.

A meta-análise dos estudos não é possível porque os mesmos não apresentaram ou mensuraram as variáveis da mesma maneira. Cousins e Leithwood compensaram esse problema por intermédio da criação de um índice que balanceava o número de achados positivos, negativos e não-significantes para cada característica, para criar o índice de “relação de prevalência”. Baseado nesse índice, eles concluíram que a qualidade da avaliação e as características da decisão estavam mais fortemente relacionadas com a utilização, seguidas dos achados, do comprometimento ou receptividade da avaliação e da relevância.

Esse índice fornecia os meios para comparação dos achados entre os estudos. Infelizmente, apresentar conclusões sobre que características estão relacionadas à utilização continua problemático, porque o tipo de meta-síntese é altamente afetado pelos componentes escolhidos pelos pesquisadores, e isso pode não incluir o que ocorreu de fato. Ao contrário, o presente estudo discute aqueles elementos que parecem ser mais “empiricamente suportados”- significando aqueles elementos que são mais estudados e suportados por evidências fortes da relação positiva com a utilização da avaliação - reestruturando a argumentação para discutir as características “empiricamente suportadas”, também permite a sugestão de evidências baseadas na prática que os avaliadores podem empregar para aumentar a utilização das avaliações (JOHNSON *et al.*, 2009, p. 388-389, tradução nossa).

2.10 OS MODELOS TEÓRICOS DA UTILIZAÇÃO DA AVALIAÇÃO

Conforme demonstrado nas seções anteriores, os trabalhos de Leviton e Hughes (1981) e Cousins e Leithwood (1986) e, mais recentemente, o trabalho de Johnson *et al.* (2009) sistematizam os fatores que influenciam a utilização, evidenciando a complexidade do

estudo da utilização e, conseqüentemente, a complexidade dos contextos em que se realizam as avaliações de programas e políticas, o que levou ao desenvolvimento de teorias da utilização.

Nesse sentido, o meta-modelo hipotético de Johnson (1998), apresentado a seguir, procura integrar a teoria e prática com vistas à construção de um modelo conceitual sobre as diferentes modalidades de utilização da avaliação. Posteriormente, outras teorias da utilização foram sendo desenvolvidas ou aperfeiçoadas, como a teoria das avaliações influentes de Mark e Henry e a teoria das avaliações focadas na utilização de Patton.

2.10.1 O meta-modelo de utilização proposto por Johnson (1998)

R. Burke Johnson (1998) propõe a integração da literatura e da teoria sobre utilização da avaliação por meio de um meta-modelo hipotético de utilização, formulado com base nos modelos implícitos e explícitos de utilização desenvolvidos por diversos autores. Nos modelos implícitos, a utilização não é diretamente descrita pelo autor, já os modelos explícitos de utilização foram construídos e testados pelos autores.

O modelo analítico adotado pelo autor compreende a identificação da tipologia de utilização, bem como dos modelos de processos de utilização da avaliação disponíveis na literatura, seus fundamentos e variáveis explicativas. Com base nessa revisão, são apresentados os fatores que facilitam a utilização das avaliações, que subsidiaram a construção do modelo proposto pelo autor (FIGURA 2).

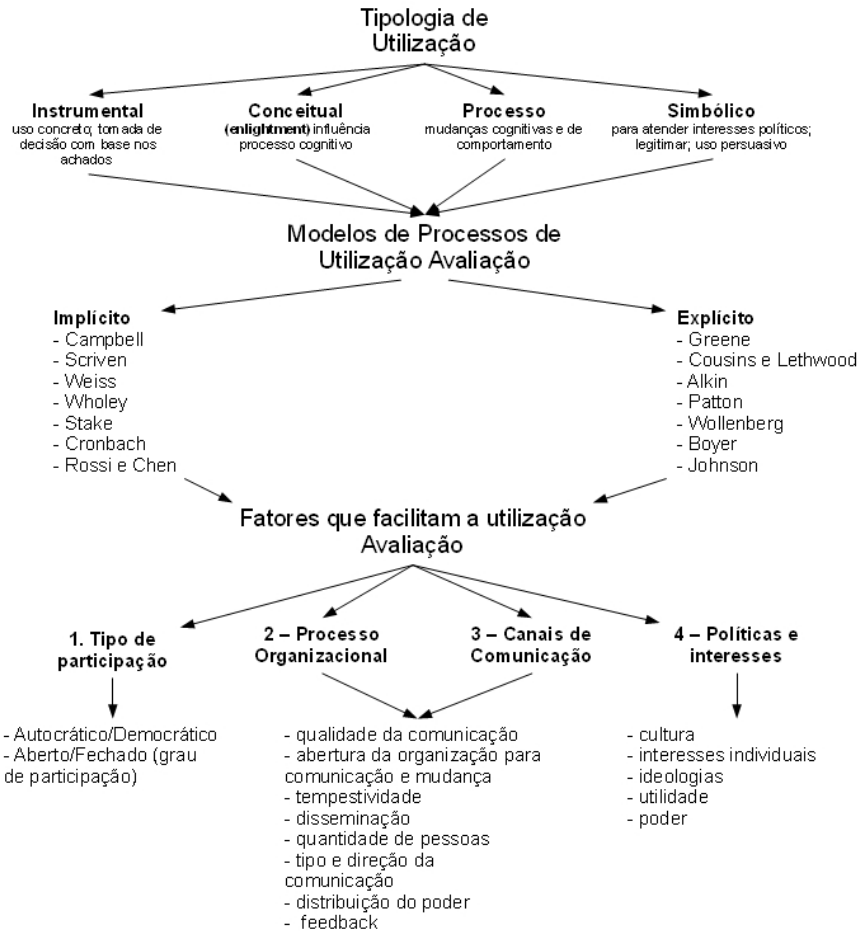


FIGURA 2 - MODELO ANALÍTICO DE JOHNSON (1998)

Os fatores que facilitam a utilização da avaliação, identificados na FIGURA 3, considerados chave pelo autor, dizem respeito ao tipo e grau de participação – tanto dos avaliadores, como dos demais envolvidos na avaliação –, ao processo organizacional e aos canais de comunicação, que podem ser mensurados pelas características e qualidade da comunicação, pela abertura da organização para comunicação e mudança, assim como pela tempestividade e abrangência da disseminação das informações relativas a avaliação. Outro fator chave destacado, refere-se ao número de pessoas e organizações responsáveis pela disseminação e discussão dos achados da avaliação, bem como a direção da comunicação (vertical, horizontal, diagonal), frequência e o *feedback*. Além disso, a utilização da avaliação é afetada pela distribuição de poder dentro da organização.

Assim como Cousins e Leithwood, Johnson destaca os fatores relacionados com o contexto político e os interesses próprios dos tomadores de decisão, como importantes fatores para a utilização.

Para Johnson, portanto, a utilização da avaliação é um processo contínuo e difuso, interdependente do contexto local e das dimensões organizacional e política. A participação dos *stakeholders* do programa é essencial, assim como a disseminação, comunicação e o

feedback de informações e resultados, tanto para os usuários, como para os avaliadores, durante e depois da avaliação. Esses fatores contribuem para aumentar a utilização, a relevância do trabalho, bem como favorecem as modificações necessárias nos programas e na apropriação dos resultados pelos atores. Desta forma, avaliadores, gerentes e outros *stakeholders*-chave deveriam, de forma colaborativa, desenvolver mecanismos para promover o aumento da qualidade da participação, disseminação, utilização e aprendizado organizacional.

Assim, um meta-modelo teórico do processo de utilização (FIGURA 3) foi proposto pelo autor, onde um conceito multidimensional de utilização foi desenvolvido envolvendo a utilização cognitiva (baseada na avaliação participativa), comportamental (instrumental) e o aprendizado organizacional.

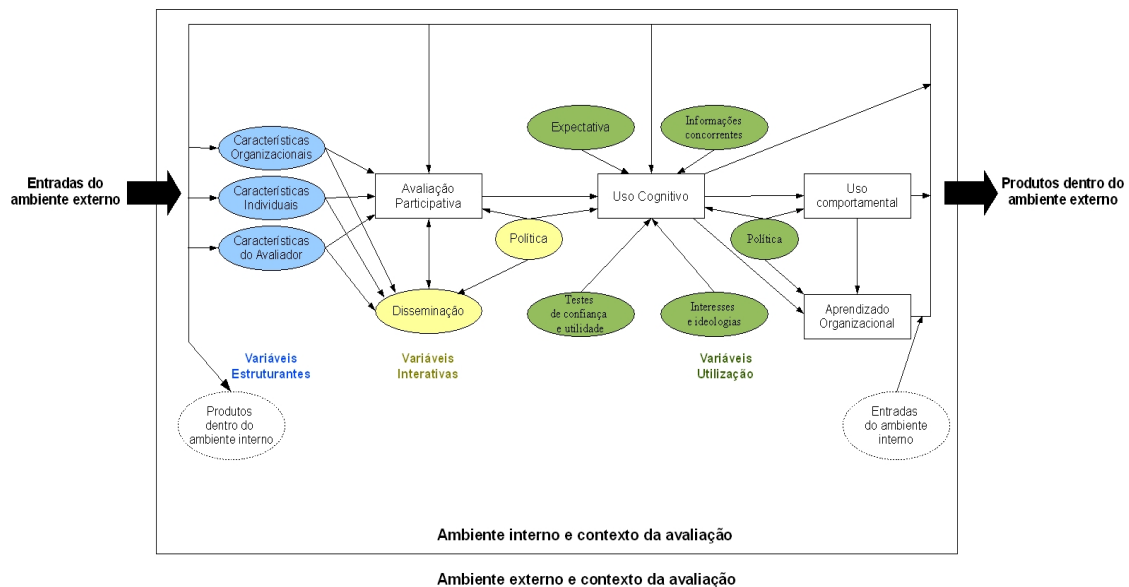


FIGURA 3 - MODELO TEÓRICO DE UTILIZAÇÃO DA AVALIAÇÃO DE JOHNSON (1998, P. 104, TRADUÇÃO NOSSA)

Conforme podemos visualizar na FIGURA 3, a utilização é interpretada como um processo contínuo e interdependente, cuja participação dos diversos *stakeholders* é moldada pelas características organizacionais, dos individuais e do avaliador e influencia todo o fluxo da avaliação, cujos canais são a comunicação. A utilização cognitiva precede a comportamental e o aprendizado organizacional. A utilização cognitiva é interpretada como o núcleo do processo, pois é onde se formam as opiniões, valores e atitudes sobre o programa, como resultado da avaliação e da participação no seu processo de execução. É também no nível cognitivo que se desenvolvem as habilidades que afetaram o comportamento, que se traduzirá nas ações (utilização instrumental), mas também poderá incluir a utilização simbólica.

Quanto ao aprendizado organizacional, que seria outra modalidade de utilização, Johnson ressalta que,

O aprendizado organizacional é uma teoria interdisciplinar de mudança, que integra ideias de diferentes teorias, mas não estão limitadas a elas, como da teoria do aprendizado, comportamento organizacional, teoria organizacional, desenvolvimento organizacional, construtivismo, teoria da complexidade e ciência cognitiva (Coveny e Highfield, 1995; Guba e Lincoln, 1989; Mainzer, 1994; Senge, 1990; Waldrop, 1992). O foco dessa teoria está em como as pessoas agem no sistema de aprendizado dinâmico, como criam e compreendem novas ideias, como se adaptam às situações de mudanças constantes, e como novos procedimentos e estratégias são incorporadas à cultura organizacional, estrutura, políticas e à mente humana. É uma teoria de sistemas dinâmicos e complexos (COVENY E HIGHFIELD, 1995; SENGE, 1990). (JOHNSON, 1998, p. 103, tradução nossa).

O aprendizado organizacional envolve tanto mudanças nas pessoas como na estrutura, nos procedimentos e cultura organizacionais e deve ser considerado pelo avaliador, não como um efeito, mas como um meio para se compreender e promover a continuidade das organizações, a longo prazo, em ambientes dinâmicos complexos.

O meta-modelo hipotético de Johnson (FIGURA 3) é muito similar ao de Cousins e Leithwood (FIGURA 1), onde a utilização cognitiva pode ser associada ao processamento de informações, à utilização para decisão, ao uso comportamental, e ambos consideram o aprendizado como um tipo de utilização. Ocorre que Cousins e Leithwood restringiram o aprendizado ao tomador de decisão e Johnson amplia a utilização do nível indivíduo para o nível organizacional.

O que os diferencia é justamente o propósito da pesquisa, Cousins e Leithwood estavam preocupados em mensurar a importância dos fatores para a utilização e, Johnson, em construir um modelo teórico da utilização.

Segundo Blake e Ottoson (2009),

Johnson apresenta um meta-modelo de utilização que sugere que a utilização da avaliação ocorre em um ambiente interno inserido no contexto externo. As variáveis associadas com a utilização são agrupadas em três categorias principais: estruturantes, interativas ou psicossocial e utilização da avaliação. O ciclo de feedback do modelo sugere que a utilização da avaliação é contínua, evolui e muda a forma com o tempo, como também é um processo interativo, multi-stakeholder, de reflexão e adaptação do programa (JOHNSON, 1998). Johnson argumenta que as ideias da teoria complexa, aprendizado organizacional e desenho organizacional aumentam a utilização e a efetividade das avaliações (BLAKE e OTTOSON, 2009, p. 25, tradução nossa).

Por outro lado, Mark e Henry, como veremos a seguir, introduziu no modelo de Cousins (2003) um novo elemento para compreensão do fenômeno utilização das avaliações, que são os mecanismos de utilização.

2.10.2 Avaliações Influentes de Mark e Henry (2004)

Para compreensão das dimensões da utilização, Mark & Henry (2004) apresentaram uma analogia entre as teorias atuais da utilização e as teorias do programa, utilizando o modelo lógico elaborado por Cousins (2003), na revisão do uso no contexto na avaliação participativa. O processo avaliativo é apresentado seguindo a estrutura do modelo lógico, no qual os insumos (entradas) são o contexto da avaliação e do conjunto de variáveis dentro do qual a avaliação é realizada (decisório/político). O contexto da avaliação inclui os recursos humanos e outros alocados na avaliação, enquanto que o contexto decisório/político engloba os aspectos culturais, políticos e informativos da organização responsável pela implementação do programa ou da política, onde tem início a avaliação. Essas entradas (*inputs*) orientam as atividades avaliativas, como a identificação, seleção e envolvimento dos *stakeholders*, coleta e análise de dados e a elaboração e disseminação do relatório. O produto dessas atividades (*output*) é o conhecimento produzido pela avaliação, que deve seguir alguns parâmetros (atributos), como responsividade, credibilidade, sofisticação, comunicação e tempestividade. Na medida em que o conhecimento produzido observa esses parâmetros, a utilização desse conhecimento é favorecida, por intermédio da influência que exerce tanto na dimensão do processo avaliativo (uso no processo), como nas dimensões da utilização dos achados (FIGURA 4).

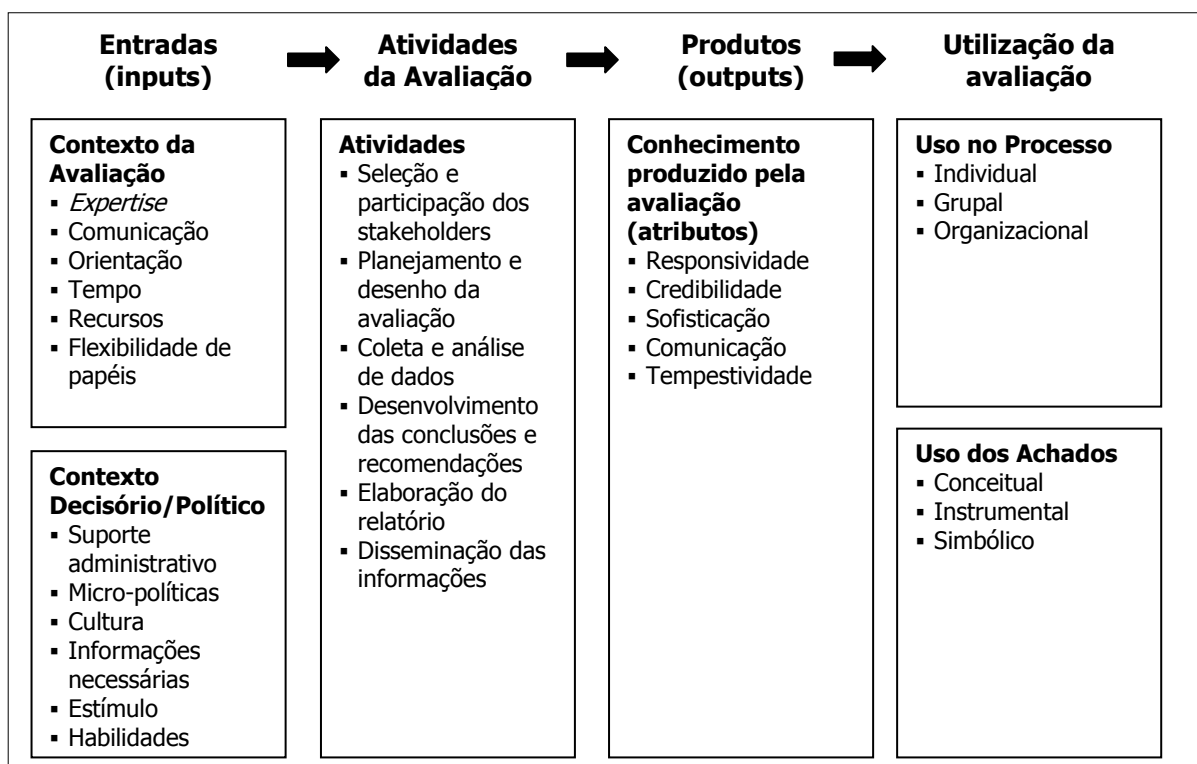


FIGURA 4 - MODELO LÓGICO DA AVALIAÇÃO

FONTE: COUSINS (2003) APUD MARK & HENRY (2004, P. 38, TRADUÇÃO NOSSA)

No entanto, Mark e Henry (2004) consideraram que existia uma lacuna, nas teorias da utilização, entre o conhecimento produzido (informações geradas - *outputs*) e os tipos de utilização, que decorre da ausência de definição dos mecanismos subjacentes ao processo avaliativo, por intermédio dos quais os seus efeitos (consequências) seriam produzidos.

Em que pese a multiplicidade de dimensões contemplada na literatura, o conceito de utilização apresenta significados conflitantes, como, por exemplo, o uso no processo pode estimular tanto o uso instrumental, por meio de ações, como o uso conceitual, mediante a compreensão. Isso se dá em razão do entendimento que se tem do que seria a utilização. Em geral, a utilização é tratada de forma descritiva, em termos de potenciais consequências produzidas pela avaliação (benefícios esperados), ou de forma normativa, mediante orientações/recomendações formuladas pelas avaliações que deveriam gerar determinadas consequências (efeitos). Ademais, Mark e Henry (2004) acrescentam que essa falta de precisão do conceito de utilização decorre, em parte, das diferentes posições defendidas pelos teóricos, alguns entendem que a utilização está restrita ao contexto e período contemplados na avaliação (como exemplos citam: ALKIN, 2003; HOFSTETTER and ALKIN, 2003 *apud* MARK e HENRY, 2004), enquanto que, para outros, o domínio da avaliação é amplo, incluindo mudanças que podem ocorrer fora do seu contexto original.

Não obstante o conceito de avaliação influente proposto por Kirkhart (2000) busque conciliar esses posicionamentos, contemplando as questões relativas a quando e onde, Mark e Henry (2004) sugerem que o foco da teoria da avaliação influente deveria ser nos resultados da avaliação (*outcomes*), naqueles relacionados às mudanças de atitudes em relação à prestação de serviços ou à política, que propiciem o redirecionamento para o objetivo maior da avaliação, a melhoria social. Para tanto, propõem um conjunto de mecanismos mediante os quais se poderia mensurar a influência da avaliação nos resultados alcançados. Esses mecanismos são classificados em quatro processos: influência geral, cognitivo e afetivo (ou relacionado à atitude), motivacional e comportamental. Para cada um deles, as consequências são analisadas em três níveis: individual, interpessoal e coletivo. Por intermédio desses mecanismos, pode-se, de forma mais objetiva, descrever como ocorre a utilização nas dimensões tradicionalmente apresentadas, especialmente quanto ao uso instrumental, que estaria relacionado aos mecanismos comportamentais; e ao uso conceitual, que apesar de suas características, ou seja, apesar de mensurar a apreensão/compreensão de novos conceitos/ideias pelo indivíduo ou pela organização, pode ser percebido pelos mecanismos cognitivo/afetivos. Quanto ao uso simbólico, os autores consideram que essa dimensão da utilização poderia ser percebida no nível interpessoal e coletivo, por meio da justificativa e do

ritualismo. No que diz respeito ao uso ritualístico no nível coletivo, ressaltam que esse expressa a tomada de decisão racional (QUADRO 14).

Com relação ao uso no processo, Mark e Henry apontam que não existe uma correspondência direta com os níveis de análise apresentados no QUADRO 14, considerando que o uso no processo significa que algumas mudanças aconteceram em razão do processo de avaliação e não em função dos achados, podendo ser acompanhados muito mais pelas atividades da avaliação, do que pelos seus produtos (MARK e HENRY, 2004, p. 44).

Tipos de Processo/Resultado	Níveis de Análise		
	Individual	Interpessoal	Coletivo
Influência Geral	Elaboração (formação de opinião)	Justificativa	Ritualismo (decisão racional)
	Heurística	Persuasão	Receptividade legislativa
	Priming (1)	Agente de mudança	Formação de coalizão
	Novas habilidades	Influência da opinião da minoria	Projetos de Lei Definição de padrões Considerações políticas
Cognitivo e afetivo (uso conceitual)	Relevância	Normas descritivas locais (Local descriptive norms)	Definição da agenda (<i>agenda setting</i>)
	Opinião positiva/negativa		Aprendizado orientado para a política
Motivacional	Objetivos e aspirações pessoais	Normas restritivas (<i>injunctive norms</i>)	Estrutura de incentivos
		Recompensas sociais	Pressão do mercado
		Intercâmbio	
Comportamental (uso instrumental)	Desempenho de novas habilidades	Mudança nas atitudes coletivas	Continuação, finalização ou alteração de programas
	Mudança nas atitudes individuais		Mudanças na política
			Disseminação

QUADRO 14 - MODELO DE MECANISMOS ALTERNATIVOS QUE PODEM MEDIAR A AVALIAÇÃO INFLUENTE

FONTE: (MARK E HENRY, 2004, P. 41, tradução nossa)

NOTA: (1) Quando um conceito é utilizado para influenciar o julgamento de outros, inédito, prioritário

A partir desses mecanismos, poder-se-ia construir os caminhos (*pathways*) por intermédio dos quais a influência é exercida, favorecendo uma análise mais percuciente da utilização do conhecimento produzido pela avaliação ao longo de todo o seu processo, e não apenas no seu estado final.

2.10.3 O modelo de Patton - Avaliações focadas na utilização

Michael Quinn Patton, já citado anteriormente, quando do debate com Carol Weiss, é o autor da teoria da avaliação focada na utilização, cuja primeira edição de sua obra foi em 1978 e, nos dias atuais, encontra-se na sua quarta edição. Ele tem sido uma referência nas

pesquisas sobre a utilização, em função da abordagem de sua teoria, que ao longo dos anos foi incorporando novas dimensões de análise, conforme podemos perceber das palavras a seguir, compiladas do prefácio da última edição de seu livro:

O tom e o conteúdo desta nova edição têm sido influenciados pelo fato de que a avaliação focada na utilização tem mais de 30 anos. A primeira edição, publicada em 1978 e baseada na pesquisa iniciada em 1975, tinha o tom de um bebê chorão, porque ninguém parecia estar prestando atenção. A segunda edição, publicada em 1986, estabeleceu *intenções de uso para os interessados no uso* como regra central da avaliação focada na utilização e adicionou exemplos para esclarecer pontos de incerteza e contenção. A partir da terceira edição, a premissa básica da avaliação focada na utilização passou a ser amplamente aceita pelos avaliadores profissionais. Joe Wholey, Harry Hatry e Kathryn Newcomer (1994) observaram no *Handbook of Practical Program Evaluation* que “nos anos recentes o lema da avaliação profissional tem sido a avaliação focada na utilização” (p. 5). Nesse contexto, a terceira edição do livro deu mais atenção aos detalhes e nuances da prática e abriu novo território para o *uso no processo*, examinando os impactos da avaliação para além dos achados, verificando como os stakeholders foram afetados por estarem envolvidos no processo de avaliação. O conceito de *uso no processo* chamou a atenção para a observação de que aqueles que se envolviam na avaliação sempre experimentavam mudanças na forma de pensar, agir e se comportar que podem afetar a forma como o programa e as organizações conduzem seus negócios. Na última década, nosso entendimento desses efeitos tem sido aprofundado consistentemente e essa nova edição revisa as últimas evidências e fornece mais diretrizes para fortalecer a capacidade avaliativa nas organizações por meio da introdução do pensamento avaliativo dentro da cultura organizacional. (PATTON, 2008, p. xvi-xvii, tradução nossa).

A avaliação focada na utilização, conforme definida por Patton,

é a avaliação feita para e com vistas a atender intenções específicas dos usuários imediatos, para utilização nesses fins. A avaliação focada na utilização tem como premissa que serão julgadas pela sua utilidade e utilização; portanto, os avaliadores devem facilitar o processo de avaliação e planejar cada avaliação observando cuidadosamente como as coisas são feitas, do início ao fim, e de que forma afetarão a sua utilização. A utilização diz respeito a como as pessoas reais, no mundo real, aplicam os achados das avaliações e experimentam o processo de avaliação. Desta forma, o foco da avaliação voltado para a utilização deve ser nas intenções de uso para os interessados na utilização (*intended use by intended users*). (PATTON, 2008, p. 37, tradução nossa).

Essa abordagem é pessoal e situacional, na medida em que a avaliação é construída a partir da identificação dos “interessados na utilização”, indivíduo ou grupo de pessoas que tem interesse nos achados produzidos pela avaliação, que serão aplicados para fins específicos, que determinam a seleção do modelo, método e teoria melhor plasmados a uma situação em particular.

Para Patton, existem diferenças entre a avaliação focada na utilização e a avaliação influente de Mark e Henry, considerando que:

a ênfase da abordagem focada na utilização é na intencionalidade e direcionamento que propositalmente favorecem a utilização. Ao contrário, a avaliação influente

ênfatiza os aspectos indiretos sobre os efeitos da avaliação no tempo e fora do programa avaliado, que estão além do controle do avaliador. Os avaliadores focados na utilização, por outro lado, podem conduzir seus trabalhos de forma a promover a sua utilização, especialmente quando observam os propósitos que deram origem à avaliação. (PATTON, 2008, p. 110, tradução nossa).

Os propósitos estabelecidos para a avaliação direcionam os diferentes tipos de utilização e têm implicações em todos os aspectos da avaliação: desenho, indicadores, análise, interpretação, elaboração do relatório e disseminação, como também com os critérios para julgamento da qualidade. O autor apresenta uma tipologia de utilização segundo as distintas finalidades do uso (QUADRO 15).

Finalidades do Uso	Tipos de Uso
Direta	Uso Instrumental Uso Conceitual Uso no Processo
Longo prazo - influência incremental	Influente <i>Enlightenment</i>
Político	Uso simbólico Uso para legitimar Uso persuasivo Uso impositivo Uso mecânico
Mau uso	Mau uso doloso (<i>mischievous misuse</i>) Mau uso culposo (<i>inadvertent misuse</i>) Abuso
Não uso	Não uso por inconsistências na avaliação (<i>nonuse due to misevaluation</i>) Não uso político (<i>political nonuse</i>) Não uso agressivo (<i>agressive nonuse</i>)
Efeitos não-intencionais	Uso não-intencional

QUADRO 15 - DISTINÇÕES DE USO

FONTE - PATTON (2008, p. 112-113, tradução nossa)

Patton (2008) estrutura as finalidades do uso a partir das distinções clássicas, indicando aquelas cujos achados das avaliações são utilizados diretamente para subsidiar a tomada de decisão, para o aperfeiçoamento de programas/políticas ou para solução de problemas (uso instrumental); quando são utilizadas para influenciar pessoas tem-se o uso conceitual; e quando promovem o aprendizado e mudanças durante o processo de avaliação, serve ao uso no processo. No entanto, quando se tem a intenção de produzir efeitos para além do período avaliativo e de seu contexto, tem-se a avaliação influente, assim como o *enlightenment*. Nos casos em que as avaliações objetivam oferecer subsídios para o uso político, tem-se o uso simbólico, assim como o uso para legitimar decisões/posições, que seria uma variação do uso simbólico, e o uso persuasivo. Patton inclui nesta finalidade de utilização, o uso impositivo, apresentado por Weiss; e o uso mecânico ou de conformidade. Nesse tipo de avaliação, a motivação é a regularidade (legalidade) e as medidas corretivas decorrentes dos achados são mecânicas (conformidade legal).

Patton insere nas finalidades do uso, por mais contraditórias que possam parecer, aquelas modalidades que poderiam ser consideradas não desejadas e que têm desafiado os pesquisadores a compreendê-las e defini-las. Trata-se do mau uso, o abuso, o não uso e o uso não intencional.

2.10.4 O mau uso, o abuso, o não uso e o uso não intencional

Segundo Shulha e Cousins (1997), os primeiros notáveis a conceituar as diferenças entre essas questões foram Alkin e Coyle (1988), distinguindo entre o não-uso-justificado das avaliações, do mau uso, da avaliação inconsistente e do abuso.

Para Patton (2008) o mau uso inclui a supressão intencional, a má apresentação ou a desproporcionalidade do uso dos achados para influenciar opiniões e decisões. O mau uso culposo (*inadvertent misuse*), também denominado “mau uso equivocado”, ocorre quando aqueles que utilizam os achados não têm a competência e o conhecimento necessários para interpretá-los adequadamente. Quanto ao abuso, este se verifica quando muita ênfase é dada aos achados, como por exemplo, quando os resultados de uma avaliação deficiente, em termos de evidências, são considerados como definitivos e utilizados de forma impositiva. Já o não uso por inconsistências na avaliação, ou o não uso justificado, refere-se à não utilização em razão de deficiências nas avaliações, como evidências fracas, relatório intempestivo, baixo desempenho do avaliador, dentre outras. No que diz respeito ao não uso político, Patton aponta que esse ocorre quando os achados são ignorados em razão de conflitos com os valores, preferências, ou predisposições dos potenciais usuários. O não uso agressivo ou resistência calculada se refere àquelas situações onde a avaliação ou o avaliador são atacados e a utilização descaracterizada, porque os seus resultados conflitam com ou levantam questões sobre uma determinada posição preferida. O uso-não-intencional é aquele em que os achados foram aplicados de forma não prevista e não planejada.

2.11 A NÃO UTILIZAÇÃO DOS RESULTADOS DAS AVALIAÇÕES NO SETOR PÚBLICO

Apresentadas as pesquisas e teorias que investigaram a utilização das avaliações, entendemos pertinente inserir neste capítulo, algumas evidências de pesquisas já realizadas no setor governamental, que demonstram que, em que pese a ampla disseminação da função avaliativa nos governos, a utilização dos seus resultados ainda é deficiente, em razão,

principalmente, da inadequação das premissas e da abordagem metodológica que não levam devidamente em conta a questão da utilização.

As reformas gerenciais do setor público nas últimas décadas têm sido orientadas pela abordagem voltada para o negócio, para a gestão por resultados, com ênfase na melhoria da qualidade dos serviços públicos e na responsabilização por desempenho. Para assegurar que os resultados sejam atingidos, a função avaliação tem um papel estratégico, por intermédio da qual se realiza a mensuração e o monitoramento dos resultados, mediante o fornecimento de informações e dados sobre o grau de alcance dos objetivos definidos para programas e políticas públicas.

A avaliação de programas é o instrumento chave dos modelos gerenciais de gestão governamental. Por esta razão, os sistemas de monitoramento e avaliação foram amplamente disseminados. Assim, compreender a natureza dessas avaliações e a forma como são realizadas é fundamental para demonstrar porque não são devidamente utilizadas para a promoção das melhorias necessárias para o aperfeiçoamento da administração pública.

A abordagem clássica de avaliação de programas governamentais, segundo Radin (2006), é composta por três etapas:

- i) identificação dos objetivos de médio e longo prazo definidos para os programas, assim como dos principais resultados esperados, tanto em termos de produtos (*outputs*⁵), quanto de impactos (*outcomes*⁶);
- ii) construção de indicadores para aferição do desempenho; e,
- iii) coleta, verificação e análise dos dados com vistas a avaliar o nível de cumprimento dos objetivos estabelecidos.

Para Radin (2006), no entanto, a avaliação de programas governamentais envolve três paradoxos:

- i) espera-se que os resultados das avaliações possam contribuir para melhorar a alocação dos recursos no orçamento;
- ii) a ênfase é na mensuração do impacto, sem que seja considerada devidamente a complexidade e dificuldade de se medir os resultados de programas que envolvem diversos atores com diferentes agendas e conflitos de valores dentro de um processo decisório fragmentado; e,

⁵ Output - produtos e serviços disponibilizados. Constituem-se no produto final de um processo de trabalho (Radin, 2006, p. 15)

⁶ Outcome - um evento, ocorrência ou condição que está fora das fronteiras do processo de execução da atividade ou programa, mas que emerge dos seus objetivos. Resultado final esperado (impacto).

- iii) a relação de confiança entre os burocratas e os cidadãos é conflituosa, em que pese o reconhecimento da sociedade de que para o alcance dos objetivos dos programas, depende-se da burocracia.

Esses paradoxos decorrem, fundamentalmente, da tensão entre os valores políticos e administrativos, cuja ênfase é na eficiência em detrimento de outros valores que emergem do sistema democrático e constantemente se confrontam com as normas de eficiência, como por exemplo, a equidade; como também na adoção das abordagens *top-down*, na confiança nos modelos de avaliação *one-size-fits-all*, que acabam por dificultar o uso das informações por eles produzidas.

Os modelos tradicionais de avaliação adotam como premissas que:

os objetivos dos programas estão claramente definidos, assim como as responsabilidades de cada um dos atores envolvidos no processo de execução; que os impactos podem ser especificados independentemente das entradas (inputs), dos processos e dos produtos (outputs), como também que podem ser mensurados, quantificados e controlados. Além disso, que os dados existem, são claros e acurados e que as avaliações de desempenho podem ser entregues para um agente com autoridade para responder pelos seus resultados (RADIN, 2006, p. 19, tradução nossa).

Fesler e Kettl *apud* Radin (2006) consideram que cinco questões afetam os impactos esperados dos programas:

- i) a incerteza que acompanha os programas;
- ii) os recursos necessários à sua execução;
- iii) as características das organizações que determinam como as burocracias reagem aos problemas;
- iv) a liderança que orienta a burocracia na solução dos problemas;
- v) a crescente interdependência entre os diferentes níveis de governo e entre esse e as entidades privadas e não-governamentais. (RADIN, 2006, p. 39, tradução nossa).

Muitos programas apresentam uma complexa combinação de objetivos (múltiplos objetivos) que visam produzir resultados eficientes, efetivos e equitativos (múltiplos valores) e usualmente os gestores públicos são desafiados a desenvolver formas de viabilizar a implementação de programas que apresentam objetivos conflitantes em relação aos resultados que se espera alcançar.

Em vista dessa complexidade, os gestores de programas preferem mensurar o desempenho com foco nos processos, utilizando indicadores de eficiência, pois medir a efetividade, o impacto, requer consistência na formulação dos objetivos dos programas e tempo para que se produzam os resultados desejados. Há que se ressaltar, ainda, que

efetividade contempla várias dimensões, que não se limitam às de natureza quantitativa, mas também a aspectos qualitativos relacionados a mudanças comportamentais, satisfação e cognição. Alguns programas têm por objetivo atingir resultados simbólicos e é extremamente difícil caracterizar e mensurar o simbólico.

O que se observa é que existe conflito de valores e interesses entre os diversos atores que participam do processo de avaliação e daqueles que deveriam fazer uso dessas informações.

Radin (2006, p. 184 e 185) considera que o modelo de avaliação de programas governamentais foi construído sob pressupostos falsos ou irrealistas, segundo os quais:

- i) a informação está sempre disponível;
- ii) a informação é neutra;
- iii) sabe-se o que está sendo mensurado;
- iv) pode-se definir a relação de causa e efeito dos programas;
- v) a linha de base da informação (padrões) existe e está disponível;
- vi) quase todas as atividades podem ser mensuradas e quantificadas.

No entanto, a informação, conforme expresso por Carol Weiss *apud* Radin, “deve ser vista não como uma iniciativa independente, mas como um recurso que tem ampla interação com a abordagem ideológica (filosofia, princípios, valores, orientação política, etc.), como também com poder, reputação e recompensas financeiras⁷.”(Radin, 2006, p. 188, tradução nossa).

Weiss (1998) ressalta que existe uma diversidade de usuários dos resultados das avaliações (*stakeholders*), cada qual com seus próprios interesses e ideias, seus próprios questionamentos e preocupações e formas de implementar os seus resultados. Portanto, a leitura da avaliação deve ser feita observando-se o seu propósito (para que fim) e para quem foi realizada.

Pode-se observar, portanto, que a premissa da utilização automática dos resultados das avaliações para a melhoria, tanto dos programas de governo, como da alocação dos recursos no orçamento, não tem se confirmado, conforme demonstram os resultados das pesquisas e trabalhos realizados por Radin (2006) e Moynihan (2008), sendo este o ponto principal para se repensar a forma produção dessas informações.

⁷ Tradução livre adotada no seminário sobre a obra *Challenging the Performance Movement* de Beryl Radin em cumprimento aos requisitos da disciplina Tópicos Especiais em Administração Pública Contemporânea, PPGA/UnB, Brasília, junho/2008.

2.11.1 Para quem importa os resultados das avaliações?

Segundo Wildavsky *apud* Moynihan (2008), “a produção e o uso das avaliações são uma forma de *interação social* caracterizada pela participação de atores com capacidade discricionária, interesses e perspectivas diferentes, que buscam o acordo por meio de barganhas, que não necessariamente é uma decisão correta.” (MOYNIHAN, 2008, p. 95, tradução nossa).

A informação não é estática, é criada para produzir efeitos sobre outros atores e o seu significado dependerá da interação entre esses atores.

Para compreender como as avaliações são utilizadas no âmbito do processo orçamentário na gestão por resultado, Moynihan (2008) utilizou a teoria do modelo diálogo interativo (*interactive dialogue model*), objetivando identificar, especificamente, que informações estão sendo apresentadas, por quem, sob que condições, que significados esperam produzir e a quem afeta.

O termo diálogo interativo sugere que a avaliação é parte da conversação política entre atores e depende da posição e do papel de cada um, assim como do nível de poder e de recursos envolvidos.

As premissas para integrar as avaliações ao orçamento tendem a assumir que a informação é objetiva, padronizada, indicativa do desempenho atual, compreendida de forma consistente e propícia para formação do consenso sobre os resultados do programa e os recursos necessários. No entanto, nas pesquisas realizadas por Moynihan (2008) estas premissas não se confirmaram. Demonstraram, ainda, que essas informações não são utilizadas de forma sistemática e, geralmente, são apresentadas como evidências para defender determinadas posições consideradas importantes por um dos atores.

Por outro lado, a habilidade para persuadir depende da boa comunicação, das rotinas de diálogo. Portanto, é importante que existam rotinas que viabilizem a comunicação. No processo orçamentário, no entanto, esses eventos são limitados. O ciclo orçamentário é muito curto, e os tomadores de decisão recebem muitas informações para serem processadas em pouco tempo, desta forma, considerando que diferentes atores podem chegar a resultados plausíveis divergentes, utilizando como base o mesmo dado, é quase impossível que consigam emitir um julgamento consensual sobre as alocações orçamentárias com base nas avaliações realizadas. Além disso, os incrementalistas sugerem que as limitações cognitivas eliminam o papel da informação de desempenho do debate.

O modelo diálogo interativo sugere que o aprendizado com base nas avaliações é maior dentro das agências do que em outros ambientes, considerando que existe uma relativa homogeneidade no grupo dos tomadores de decisão que compartilham objetivos similares. No estudo, Moynihan (2008) constatou que os gestores utilizam as rotinas do diálogo para melhorar a capacidade organizacional e reestruturar os objetivos básicos da organização.

Em síntese, podemos destacar três variáveis estratégicas no modelo diálogo interativo proposto por Moynihan: as relativas à avaliação: Quem produziu? Para qual finalidade? Em qual contexto? Aquelas referentes à capacidade institucional, e a terceira variável diz respeito ao aprendizado e sua disseminação (comunicação) (MOYNIHAN 2008).

Esses elementos nos remetem à estratégia proposta por Kettel (2002), para transformação da governança no setor público, fazendo-se refletir sobre a necessidade de se buscar a convergência dos próprios estudos da administração pública, na medida em que a causa dos problemas reforçam a constatação de que, para o alcance dos fins, necessariamente tem de se estruturar adequadamente os meios: coordenação, cooperação e comunicação.

2.11.2 O que dizem as avaliações de programas governamentais?

Radín (2006) ressalta alguns paradoxos que a experiência com a ampla disseminação das atividades avaliativas, na última década do século XX, tem evidenciado:

- i) retórica ambígua - ênfase em resultados imensuráveis;
- ii) estratégia *on-size-fits-all* e a estratégia de resposta individualizada para programas e políticas específicas; e
- iii) expectativa de que avaliação seja uma panacéia na solução dos problemas.

As avaliações governamentais esperam atingir múltiplos objetivos, cujos resultados não podem ser mensurados, muitas vezes, em razão da sua natureza simbólica. Ademais, conforme já mencionado, a informação não é neutra, carrega consigo valores e interesses pessoais e institucionais, cuja interpretação é ambígua, subjetiva e, por essas razões, sujeitas a diversos entendimentos.

Desta forma, os estudos têm demonstrado baixa relação entre os resultados das avaliações e a utilização dessas informações para a promoção de melhorias no desempenho dos programas e na *accountability* interna e externa, principalmente, em razão das questões relativas às sistemáticas e abordagens metodológicas aplicadas, que não consideram devidamente as necessidades daqueles que devem utilizar os resultados das avaliações.

Para examinar essa questão, mostra-se oportuno verificar em que medida as avaliações de programas governamentais foram concebidas considerando os fatores que favorecem a utilização dos resultados. Para tanto, é importante correlacionar os fatores apresentados nas pesquisas e teorias sobre utilização sintetizadas nesta revisão, com os padrões de avaliação, considerando que estes objetivam assegurar que as avaliações sejam efetivas para dar cumprimento aos seus propósitos.

2.12 MARCO CONCEITUAL

A partir dos estudos realizados por Leviton e Hughes (1981), Cousins e Leithwood (1986), Johnson (1998), Johnson *et al.* (2009) pode-se identificar um conjunto de fatores que favorecem a utilização da avaliação (QUADRO 16). Esses fatores podem ser agrupados em três categorias principais: aqueles relacionados ao contexto de implementação da avaliação, os relativos ao contexto político no qual as decisões pela realização e execução da avaliação foram tomadas e, a terceira categoria, diz respeito ao envolvimento dos *stakeholders* com a avaliação, identificada por Johnson *et al.* (2009) e que está intrinsecamente relacionada aos fatores identificados por Cousins e Leithwood (1986).

Categorias	Fatores que favorecem a utilização				
	Leviton e Hughes	Cousins e Leithwood	Johnson	Johnson et al	
Implementação avaliação	Relevância	Relevância		Envolvimento com relevância	
	Comunicação	Comunicação	Canais de Comunicação	Envolvimento com qualidade da comunicação	
	Processamento de informações	Qualidade da avaliação		Envolvimento com achados	
		Achados			
	Credibilidade	Credibilidade			Envolvimento com credibilidade
		Tempestividade			Competência do avaliador
Decisão/Política	Participação/Apoio dos Usuários	Compromisso/ Receptividade da avaliação	Tipo de participação	Envolvimento com compromisso ou receptividade da avaliação	
		Necessidade de informações	Políticas e interesses	Envolvimento com necessidade de informações	
		Clima político			
		Informações concorrentes	Processo Organizacional	Envolvimento com características pessoais	
		Características pessoais			
		Características da decisão			Envolvimento com características da decisão
Envolvimento stakeholders					

QUADRO 16 - FATORES QUE FAVORECEM A UTILIZAÇÃO DAS AVALIAÇÕES SEGUNDO LEVITON E HUGHES (1981), COUSINS E LEITHWOOD (1986), JOHNSON (1998) E JOHNSON ET AL. (2009)

Ao analisarmos esses fatores à luz dos padrões de avaliação de programas instituídos pelo Comitê Conjunto de Padrões para Avaliação Educacional, pode-se identificar a importância da observância dos padrões, não apenas para assegurar a credibilidade e qualidade do trabalho, mas, principalmente, para favorecer a utilização das avaliações, com vistas a assegurar o cumprimento dos seus objetivos.

O primeiro atributo da avaliação, considerado para definição dos padrões, é a utilidade, que estrutura o estudo a ser realizado, com vistas a favorecer que a avaliação disponibilizará as informações necessárias para os interessados na sua utilização. O conjunto de padrões da utilidade trata da identificação dos stakeholders (para quem se destina a avaliação), do avaliador (quem realizará o trabalho), do problema a ser investigado (escopo), dos critérios para julgamento (identificação dos valores), do relatório e da sua disseminação e, finalmente da utilização.

O primeiro padrão da utilidade diz respeito à identificação dos *stakeholders* (U1). *Stakeholders* são todos aqueles que podem ser envolvidos na avaliação ou afetados pelos seus resultados. Conhecer quem são esses atores é importante para que se possa identificar suas necessidades e, com base nelas, definir as questões a serem investigadas, o escopo do trabalho, a que se refere o padrão U3 (SANDERS, 1994).

A identificação dos *stakeholders*, sob a ótica dos fatores que influenciam a utilização, está inserida na relevância, ou seja, de quem são as necessidades que a avaliação objetiva atender? Para quem se destina a avaliação? Na pesquisa realizada por Leviton e Hughes (1981), foram identificadas as necessidades dos políticos e dos gestores e, destacou-se que as avaliações são utilizadas com mais frequência quando atendem às necessidades de um público em particular. Da mesma forma, Cousins e Leithwood (1986) identificaram o fator: necessidade de informações, que incluiu o levantamento dos diferentes públicos da avaliação. Já Johnson (1998) trata do tipo de participação, como fator que influencia a utilização.

Segundo Fleischer e Christie (2009), Cousins e Leithwood (1986) mostraram, como resultado de sua pesquisa, que as abordagens de avaliação focadas nos usuários favoreciam a utilização, porque esses foram envolvidos no processo de avaliação e estavam investidos dos benefícios da avaliação.

Após o trabalho realizado por Cousins e Leithwood (1986), a importância dos *stakeholders* para a avaliação tem sido cada vez mais significativa, conforme ressaltado por Johnson *et al.* (2009). A ampliação das abordagens de avaliação participativa tem resultado em um maior envolvimento dos *stakeholders*, sendo que, para esses autores, o envolvimento pode ser traduzido em uma nova categoria agrupadora de fatores que influenciam a utilização.

Reineke (1991, *apud* FLEISCHER e CHRISTIE, 2009) sugere que a qualidade do envolvimento é o resultado da identificação antecipada, pelo avaliador, dos *stakeholders* e o contínuo envolvimento deles enquanto se estabelece uma estrutura para o diálogo.

No entanto, a identificação dos *stakeholders* precede o envolvimento e a participação, sendo primordial e indispensável para a realização dos estudos avaliativos.

No que diz respeito à credibilidade do avaliador (padrão U2), em todas as pesquisas realizadas destaca-se o papel do avaliador como elemento necessário para garantir a credibilidade dos trabalhos, como também como indutor da utilização. Segundo Fleischer e Christie (2009, p. 160-161, tradução nossa), “muitos fatores podem influenciar o papel do avaliador, incluindo o propósito da avaliação, as necessidades de informações dos *stakeholders*, as preferências epistemológicas do avaliador e a abordagem de avaliação a ser utilizada.”

O padrão U3 trata da identificação das informações necessárias para seleção e definição do escopo do trabalho. Segundo Sanders (1994, p. 37, tradução nossa), “para ter o escopo apropriado, as avaliações precisam ser relevantes para os objetivos dos tomadores de decisão, importantes para os principais *stakeholders* e suficientemente compreensivas para suportar o julgamento do valor ou mérito”.

A definição do escopo (U3) decorre das necessidades de informações dos *stakeholders* identificados (U1), dos problemas a serem tratados no estudo avaliativo, sendo central e crítico para a utilização. Esse ponto foi exhaustivamente discutido nas pesquisas realizadas por Leviton e Hughes (1981), quanto ao reconhecimento da relevância (avaliação direcionada para as necessidades dos *stakeholders*); Cousins e Leithwood (1986) e Johnson (2009), no que diz respeito à identificação das necessidades de informações dos *stakeholders*.

Os padrões U5 e U6 tratam do relatório e da sua disseminação. Segundo Sanders (1994), o relatório pode ter várias formas, em que pese a forma escrita ser a mais utilizada. No entanto, é essencial que seja claro para o público a que se destina, tenha credibilidade e aplicação. Para que a avaliação seja útil, precisa ser compreendida. Os *stakeholders* deveriam rapidamente compreender, por meio do relatório, os propósitos da avaliação, o que foi avaliado, como a avaliação foi conduzida, que informações foram obtidas, quais as conclusões e recomendações (SANDERS, 1994, p. 49, tradução nossa).

A disseminação diz respeito à comunicação dos achados da avaliação aos interessados na sua utilização, em tempo para que a informação possa ser mais bem aproveitada. Os padrões (U5) e (U6) relacionam-se à apresentação das informações, à

qualidade da comunicação e aos canais de comunicação, que são fatores determinantes na utilização das avaliações, conforme apresentado pelos autores estudados.

O último padrão da utilidade (U7), diz respeito ao impacto da avaliação, em termos da influência da avaliação nos públicos destinatários dos trabalhos, ou seja, em que medida os resultados da avaliação influenciaram as decisões, o aprendizado e o conhecimento sobre o objeto avaliado.

Quanto aos padrões da viabilidade, quais sejam, procedimentos, viabilidade política e custo-efetividade, mostra-se relevante, para o presente estudo, destacar as questões relacionadas à viabilidade política, em razão da importância do contexto político para a utilização das avaliações.

Segundo Weiss (1993), a avaliação é um projeto racional que se realiza no contexto político. As questões políticas interceptam a avaliação em três sentidos:

1) os objetos da avaliação – políticas e programas – são fruto de decisões políticas. São propostos, definidos, debatidos e aprovados por intermédio do processo político; além disso, quando implementados estão sujeitos a pressões, conflitos de interesses, que surgem fora do jogo político;

2) como as avaliações retro-alimentam o processo de decisão, as informações por elas produzidas entram na arena política. Desta forma, os seus resultados concorrem com outras fontes de informações que têm peso no processo político;

3) a avaliação em si tem natureza política ao legitimar questões de natureza política (problemas, objetivos, estratégias, metas). Ao se utilizar os objetivos do programa como critério da avaliação, não apenas se aceita que os mesmos são “corretos”, como também se aceitam as premissas subjacentes.

Desta forma, segundo Sanders (1994, p. 71, tradução nossa), “As avaliações são politicamente viáveis na extensão em que seus propósitos podem ser alcançados com isenção e reconhecimento equitativo das pressões e ações demandadas pelos vários grupos de interesse vinculados à avaliação.”

Já os padrões da propriedade e acuracidade tratam dos fatores relacionados à credibilidade do processo de avaliação, desde a sua contratação até a apresentação dos resultados, inclusive quanto à consistência e adequação metodológica para o exame do problema investigado. A observância desses padrões é de fundamental importância para assegurar a credibilidade e qualidade das avaliações e representam fatores relevantes para a utilização, conforme apresentado nas pesquisas relacionadas na revisão de literatura e indicadas no QUADRO 17.

PADRÕES	DENOMINAÇÃO	Leviton e Hughes (1981)	Cousins e Leithwood (1986)	Johnson (1998)	Johnson (2009)
U1	IDENTIFICAÇÃO STAKEHOLDER	Relevância (para quem se destina a avaliação?)	Necessidade de informações/características e papéis tomadores de decisão/comprometimento receptividade	Tipo de participação	Envolvimento com necessidade de informações/ Envolvimento com compromisso ou receptividade
U2	CREDIBILIDADE AVALIADOR	Credibilidade (credibilidade do avaliador)	Credibilidade (credibilidade do avaliador)	Características do avaliador	Competência do avaliador/Envolvimento com credibilidade
U3	IDENTIFICAÇÃO DAS INFORMAÇÕES NECESSÁRIAS PARA SELEÇÃO E DEFINIÇÃO DO ESCOPO	Processamento de informações/ Reconhecimento da Relevância (avaliação direcionada para as necessidades dos clientes)	Necessidade de informações		Envolvimento com necessidade de informações
U4	IDENTIFICAÇÃO DOS VALORES (CRITÉRIOS)	Processamento de Informações (reconhecimento da relevância)	Credibilidade (do processo de avaliação)		Envolvimento com credibilidade (no processo)
U5	CLAREZA DO RELATÓRIO	Processamento de Informações (Apresentação das informações)	Qualidade da Comunicação	Canais de comunicação	Envolvimento com qualidade da comunicação
U6	TEMPESTIVIDADE E DISSEMINAÇÃO DO RELATÓRIO	Relevância (Tempestividade)	Tempestividade		
U7	IMPACTO DA AVALIAÇÃO (INFLUÊNCIA P/ TOMADA DE DECISÃO)	Participação e Apoio dos Usuários	Características da decisão	Processo organizacional (abertura da organização para comunicação e mudança)	Envolvimento com compromisso ou receptividade da avaliação
F1	PRODEEDIMENTOS PRÁTICOS	Credibilidade (Qualidade)	Credibilidade (credibilidade do processo de avaliação)		Envolvimento com credibilidade
F2	VIABILIDADE POLÍTICA	Participação e Apoio dos Usuários	Clima Político	Políticas e interesses	Envolvimento com compromisso ou receptividade da avaliação
F3	CUSTO-EFETIVIDADE				
P1	ORIENTAÇÃO PARA PRESTAÇÃO DOS SERVIÇOS	Processamento de Informações (reconhecimento da relevância)	Relevância		Envolvimento com necessidade de informações
P2	ACORDO FORMAL				
P3	QUESTÕES RELATIVAS A DIREITOS INDIVIDUAIS				
P4	RELAÇÕES HUMANAS				
P5	AVALIAÇÃO COMPLETA E IMPARCIAL (JUSTA)		Credibilidade (credibilidade do processo de avaliação)		Envolvimento com credibilidade
P6	APRESENTAÇÃO DOS ACHADOS		Qualidade da Comunicação	Canais de comunicação	Envolvimento com qualidade da comunicação
P7	CONFLITO DE INTERESSES	Relevância (para quem se destina a avaliação)	Informações concorrentes e comprometimento/receptividade	Políticas e interesses	Envolvimento com características da decisão
P8	RESPONSABILIDADE E FISCAL				

(continua)

(conclusão)

PADRÕES	DENOMINAÇÃO	Leviton e Hughes (1981)	Cousins e Leithwood (1986)	Johnson (1998)	Johnson (2009)
A1	DOCUMENTAÇÃO DO PROGRAMA		Credibilidade (credibilidade do Processo de avaliação)		Envolvimento com credibilidade
A2	ANÁLISE DO CONTEXTO		Clima político	Política e Interesses/Processo Organizacional	
A3	DESCRIÇÃO DOS PROPÓSITOS E PROCEDIMENTOS		Qualidade da avaliação (sofisticação metodológica)		Envolvimento com credibilidade
A4	FONTES DE INFORMAÇÕES DEFENSÁVEIS		Qualidade da avaliação (sofisticação metodológica)		Envolvimento com credibilidade
A5	INFORMAÇÕES VÁLIDAS		Qualidade da avaliação (sofisticação metodológica)		Envolvimento com credibilidade
A6	INFORMAÇÕES CONFIÁVEIS		Qualidade da avaliação (sofisticação metodológica)		Envolvimento com credibilidade
A7	INFORMAÇÕES SISTEMATIZADAS		Qualidade da avaliação (sofisticação metodológica)		Envolvimento com credibilidade
A8	ANÁLISE QUANTITATIVA DAS INFORMAÇÕES		Qualidade da avaliação (sofisticação metodológica)		Envolvimento com credibilidade
A9	ANÁLISE QUALITATIVA DAS INFORMAÇÕES		Qualidade da avaliação (sofisticação metodológica)		Envolvimento com credibilidade
A10	CONCLUSÕES JUSTIFICADAS		Achados		Envolvimento com achados
A11	RELATÓRIO IMPARCIAL	Credibilidade avaliação	Qualidade da comunicação	Canais de comunicação	Envolvimento com qualidade da comunicação
A12	META-AVALIAÇÃO				

QUADRO 17 - RELAÇÃO ENTRE OS PADRÕES DE AVALIAÇÃO DE PROGRAMAS E OS FATORES QUE INFLUENCIAM A UTILIZAÇÃO

Desta forma, o relacionamento apresentado no Quadro 17, entre os padrões de avaliação definidos pelo *Joint Committee* e os fatores que influenciam a utilização das avaliações, apresentados na revisão da literatura, explicita o marco conceitual que embasa a operacionalização deste estudo.

Os fatores que favorecem a utilização podem ser examinados a partir dos padrões de avaliação, especialmente aqueles que dizem respeito a:

- i) identificação dos principais stakeholders: aqueles atores que devem fazer uso da avaliação para promoção dos resultados esperados com a realização do trabalho;
- ii) identificação do problema a ser investigado e dos critérios/valores a serem utilizados para o julgamento dos achados, sob a perspectiva dos stakeholders identificados;
- iii) identificação das formas de comunicação e apresentação dos resultados;
- iv) identificação do contexto da implementação da avaliação e do contexto político;
- v) credibilidade do avaliador;
- vi) credibilidade e qualidade do processo de avaliação.

3 METODOLOGIA

Esta pesquisa tem por objetivo tanto a revisão conceitual quanto empírica da utilização das avaliações de programas governamentais. Portanto, visa, a partir da pesquisa bibliográfica sobre a utilização das avaliações e do levantamento documental sobre as avaliações de programas governamentais no Brasil, identificar e examinar os propósitos, a tipologia, os objetivos, as sistemáticas e metodologias empregadas, a quem se destinam e como são utilizados os resultados das avaliações, especialmente aquelas realizadas pela SPI/MPOG, SFC/CGU e TCU. Espera-se verificar em que medida as estratégias e abordagens metodológicas adotadas consideram as necessidades e prioridades daqueles atores e/ou organizações que devem utilizar os resultados e informações produzidos pelas avaliações para dar cumprimento aos objetivos propostos.

Para desenvolver a pesquisa e atender ao objetivo apresentado, propôs-se realizar, preliminarmente, pesquisa bibliográfica com vistas a identificar os principais aspectos conceituais a serem aplicados no estudo da utilização das avaliações de programas governamentais, bem como identificar os critérios a serem utilizados para a análise das avaliações selecionadas para o estudo empírico.

Esta estratégia foi necessária considerando que, no Brasil, os estudos sobre a utilização das avaliações são incipientes, demandando a revisão da literatura sobre o tema com vistas a introduzir os principais conceitos e abordagens empregadas no estudo desse fenômeno.

Desta forma, a partir da consolidação da literatura, foram apresentadas as pesquisas e teorias consideradas importantes para o estudo da utilização das avaliações, das quais foram extraídos os fatores que influenciam a utilização, conforme demonstrado no capítulo 2.

Com base nos fatores que influenciam a utilização, buscou-se identificar um instrumento que possibilitasse examinar, nas avaliações de programas selecionadas para o estudo empírico, a observância desses fatores. Foi depreendido que os padrões de avaliação de programas definidos pelo *Joint Committee on Standards for Educational Evaluation* refletiam, de forma consistente, os principais atributos da prática avaliativa e os fatores identificados.

Assim, os padrões de avaliação de programas do *Joint Committee* foram definidos como critérios para examinar as avaliações de programas governamentais no Brasil, com vistas a analisar em que medida as sistemáticas e abordagens metodológicas aplicadas

favorecem a utilização dos resultados por elas produzidos, principalmente, no que diz respeito àquelas realizadas pelas organizações públicas responsáveis pela função avaliativa.

Uma vez definidos o arcabouço conceitual e os critérios a serem aplicados no exame da utilização das avaliações no Brasil, buscou-se, por intermédio do levantamento documental na legislação relativa à administração pública federal, identificar a tipologia, os objetivos, as características e para quem se destinam as avaliações de programas. Essas informações serviram para elaborar o perfil das avaliações governamentais e classificá-las de acordo com os propósitos elencados na literatura revisada, conforme apresentado no capítulo 4.

De posse dessas informações, realizou-se o estudo das sistemáticas e abordagens metodológicas utilizadas para realização das avaliações dos programas do Plano Plurianual 2004-2007, sob a coordenação da Secretaria de Planejamento e Investimentos Estratégicos do Ministério do Planejamento, Orçamento e Gestão; bem como das avaliações de programas do Tribunal de Contas da União, realizadas pela Secretaria de Fiscalização e Avaliação de Programas Governamentais, por intermédio das auditorias operacionais; das avaliações incumbidas à Secretaria Federal de Controle Interno da Controladoria Geral da União, processadas por meio de fiscalizações e auditorias; e ainda, das avaliações decorrentes da Política de Monitoramento e Avaliação do Ministério do Desenvolvimento Social e Combate à Fome, sob a responsabilidade da Secretaria de Avaliação e Gestão da Informação.

Além disso, procedeu-se ao levantamento de informações sobre os relatórios de avaliação publicados, especialmente os relativos ao Plano Plurianual, da SPI/MPOG, das avaliações de programas do Tribunal de Contas e da Controladoria Geral da União e aqueles disponibilizados pelo Ministério do Desenvolvimento Social e Combate a Fome, com vistas a dimensionar o universo a ser estudado e a estratégia metodológica mais adequada para examinar as seguintes questões:

- i) Qual a credibilidade e qualidade das avaliações governamentais de programas?
- ii) Qual a adequação e consistência das abordagens metodológicas adotadas nessas avaliações às necessidades daqueles que farão uso das informações?
- iii) Quais os fatores e os mecanismos que influenciam a utilização dos resultados das avaliações?
- iv) Em que medida as abordagens metodológicas adotadas nas avaliações favorecem a sua utilização?

Em vista da diversidade e quantidade de informações e considerando o nível de institucionalização da função avaliativa no Brasil, delimitou-se o período de 2004 a 2007 para

o levantamento das informações necessárias à quantificação das avaliações realizadas pelos órgãos acima mencionados.

Por meio de estudo exploratório, foram realizadas entrevistas com dirigentes e técnicos da SPI/MPOG, SFC/CGU, SEPROG/TCU e SAGI/MDS, com vistas a conhecer as sistemáticas de avaliação adotadas e as abordagens metodológicas aplicadas, bem como coletar os dados e informações referentes às avaliações por eles realizadas.

Assim, com fundamento na revisão de literatura apresentada no capítulo 2 e na análise da relação entre os padrões de avaliação de programas e os fatores que influenciam a utilização (marco conceitual) e considerando, ainda, as questões a serem respondidas na pesquisa empírica, mostrou-se mais adequada a utilização dos estudos de casos múltiplos, conforme definido por Yin (2005).

Cabe destacar, no entanto, que alguns fatores comprometeram o pleno cumprimento dos objetivos desta pesquisa, principalmente, no que diz respeito à parte empírica, em razão da limitação de prazo para conclusão dos trabalhos, bem como da não disponibilização tempestiva das informações. Tais fatos impossibilitaram contemplar a análise dos dados das avaliações realizadas pela Secretaria Federal de Controle Interno – SFC/CGU e da Secretaria de Avaliação e Gestão da Informação – SAGI/MDS, muito embora as respectivas sistemáticas de avaliação tenham sido levantadas e apresentadas no capítulo que trata da institucionalização da função avaliação no Brasil (Capítulo 4).

Em que pese essas limitações, foram coletados e analisados os dados e informações de uma amostra não estatística das avaliações dos programas do PPA 2004-2007 realizadas pela Secretaria de Planejamento e Investimentos Estratégicos – SPI/MPOG – e das avaliações executadas pela Secretaria de Fiscalização e Avaliação de Programas do Tribunal de Contas da União no mesmo período. Assim, foi examinada em que medida essas avaliações guardam conformidade com os padrões de avaliação, sem a pretensão de se generalizar os achados observados, apenas de evidenciar o nível de aderência à prática avaliativa, com ênfase na utilização, com base nos padrões mencionados.

3.1 ESTUDO DE CASOS MÚLTIPLOS

Segundo Yin (2005, p. 32), “um estudo de caso é uma investigação empírica que investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e o contexto não estão claramente definidos.”

Os estudos de caso utilizam diferentes estratégias, de acordo com o propósito da pesquisa. Podem ser exploratórios, descritivos ou explanatórios. O que os diferencia são três condições: “(a) o tipo de pesquisa proposta, (b) a extensão de controle que o pesquisador tem sobre eventos comportamentais atuais e (c) o grau de enfoque em acontecimentos contemporâneos em oposição a acontecimentos históricos.”(YIN, 2005, p. 23).

Cada tipo de pesquisa empírica possui um projeto de pesquisa implícito ou explícito, que é a estrutura que permitirá conectar os dados empíricos às questões pesquisadas. Para os estudos de caso, Yin (2005) considera que esses devem apresentar cinco componentes:

- i) as questões de estudo;
- ii) as proposições, se houver,
- iii) as unidades de análise,
- iv) a lógica que une os dados às proposições,
- v) os critérios para interpretar as constatações.

Para tratar esses cinco componentes é necessário que se disponha de uma teoria relacionada à pesquisa.

Para Van Evera (1997, p. 55), os estudos de casos servem para cinco propósitos principais: testar teorias, criar teorias, identificar condições antecedentes, testar a importância das condições antecedentes e explicar casos de importância intrínseca.

No que diz respeito às inferências que podem ser feitas a partir dos estudos de caso para as teorias, Yin (2005) faz uma ressalva quanto às diferenças entre as generalizações estatísticas e analíticas, considerando um erro crasso conceber generalizações estatísticas a partir de estudos de caso.

No caso dos estudos de casos múltiplos, Yin (2005) considera que os experimentos múltiplos podem utilizar o método de generalização analítica

no qual se utiliza uma teoria previamente desenvolvida como modelo com o qual se devem comparar os resultados empíricos do estudo de caso. Se dois ou mais casos

são utilizados para sustentar a mesma teoria, pode-se solicitar uma replicação. Os resultados empíricos podem ser considerados ainda mais fortes se dois ou mais casos sustentam a mesma teoria, mas não sustentam uma teoria concorrente igualmente plausível. (YIN, 2005, p. 54)

Yin (2005) classifica os projetos de estudo de caso em 4 tipos segundo o número de casos e o número de unidades de análise. Projetos de caso único podem ser holísticos: um único caso e seu contexto; ou incorporados: um único caso, seu contexto e mais de uma unidade incorporada de análise. Da mesma forma, os projetos de casos múltiplos podem ser holísticos: mais de um caso e os respectivos contextos em que estão inseridos; e, incorporados: mais de um caso, seus contextos e para cada caso mais de uma unidade incorporada de análise.

As unidades incorporadas de análise são subunidades do caso selecionadas para investigação. Como exemplo, Yin cita o caso de uma avaliação de um programa que envolve um grande número de projetos financiados. A avaliação é o caso único e os projetos são unidades incorporadas. No entanto, um mesmo estudo pode conter mais de um caso único; nesses casos, tem-se um projeto de casos múltiplos.

“Cada caso deve servir a um propósito específico dentro do escopo global da investigação”, sendo que o importante é observar a lógica da replicação. Assim, os casos devem ser selecionados de forma a prever resultados semelhantes ou produzir resultados contrastantes (YIN, 2005, p. 68-69).

Para aplicação da abordagem da replicação aos estudos de casos múltiplos, o primeiro passo consiste no desenvolvimento da teoria e, depois de selecionados os casos e definidas as subunidades de análise, cada estudo de caso em particular consiste na verificação das convergências entre os fatos e as conclusões para cada caso, “acredita-se, assim, que as conclusões de cada caso sejam as informações que necessitam de replicação por outros casos individuais.” A abordagem da replicação dos estudos de caso, pode ser visualizada na reprodução da FIGURA 5 apresenta por Yin (YIN, 2005, p. 72).

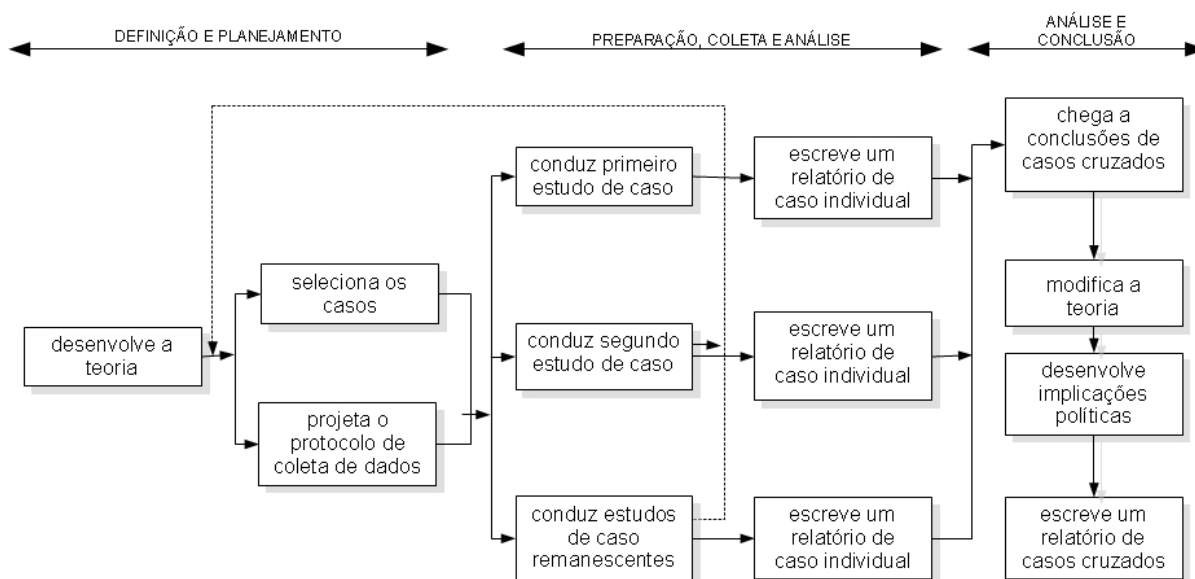


FIGURA 5 - MÉTODO DE ESTUDO DE CASO
 FONTE: COSMOS CORPORATION, YIN P. 72

O fundamento lógico para projetos de casos múltiplos deriva-se diretamente da compreensão que você tem de replicações teóricas e literais. O projeto mais simples de casos múltiplos seria a seleção de dois ou mais casos que se acredita ser replicações literais, como um conjunto de casos com resultados exemplares em relação a alguma teoria de avaliação. Selecionar esses casos exige conhecimento prévio dos resultados, com a investigação de casos múltiplos concentrando-se no como e no por que os resultados exemplares podem ter ocorrido e esperando replicações literais (ou diretas) dessas condições de caso em caso. (YIN, 2005, p. 74).

3.2 DO DESENVOLVIMENTO DA TEORIA

Para tratar os cinco componentes dos estudos de caso considerados necessários por Yin (questões de estudo, proposições, unidades de análise, lógica que une os dados às proposições e os critérios para interpretar as constatações) é necessário que se tenha uma teoria relacionada à pesquisa para ser testada ou desenvolvida.

Na revisão de literatura apresentada no capítulo 2, foram identificadas as teorias sobre a utilização da avaliação desenvolvidas por Johnson (1998), Mark e Henry (2004) e Patton (2008), além das pesquisas realizadas por Hughes e Leviton (1981), Cousins e Leithwood (1986), Johnson *et al.* (2009), que trataram muito mais da identificação das condições antecedentes e do teste da importância dessas condições para a utilização, que são os fatores que influenciam a utilização, do que propriamente de teorias. No entanto, as teorias desenvolvidas para o estudo da utilização, necessariamente, consideram os fatores que influenciam a utilização como variáveis estruturantes dessas teorias.

Desta forma, este estudo objetiva testar em que medida as sistemáticas de avaliação e as abordagens metodológicas aplicadas às avaliações governamentais de programas no Brasil observam os fatores que influenciam a utilização das avaliações apresentados nas teorias mencionadas.

A premissa desta pesquisa é que a utilização das avaliações é favorecida quando os fatores que influenciam a utilização são observados.

Para examinar a observância dos fatores que influenciam a utilização (QUADRO 16) foram utilizados como critérios de análise os padrões de avaliação de programas do Comitê Conjunto de Padrões para Avaliação Educacional, considerando a relação estabelecida entre os fatores e os padrões (marco conceitual) que foi apresentada no capítulo 2 e sintetizada no QUADRO 17.

A observância dos padrões favorece a utilização das avaliações na medida em que os fatores que influenciam a utilização estão contemplados nesses fatores.

3.3 DA SELEÇÃO DOS CASOS

Para seleção dos casos, cabe inicialmente delimitar o objeto a ser estudado nesta pesquisa, ou seja, as avaliações de programas governamentais no Brasil.

No âmbito da administração pública federal, a denominação programa – como instrumento de organização das ações de governo – é restrita aos programas integrantes do Plano Plurianual, conforme estabelecido no Decreto n. 4.052 de 13/12/2001.

Desta forma, os casos a serem selecionadas dizem respeito exclusivamente às avaliações de programas governamentais realizadas ou que estejam sob a responsabilidade dos órgãos públicos federais, cuja unidade de estudo tenha sido os programas de governo, conforme estabelecido pelo Decreto n. 4.042/2001 e definido nos normativos que orientam a elaboração dos orçamentos públicos federais.

Assim, após a realização do levantamento de informações na base de legislação, nos sítios eletrônicos do Ministério do Planejamento Orçamento e Gestão, dos órgãos de controle – CGU e TCU -, e do IPEA, constatou-se que esses órgãos realizam avaliações de programas de forma sistemática, sendo que:

- a Secretaria Federal de Controle Interno/CGU as executa por intermédio das fiscalizações e auditorias;

- o TCU, mediante as auditorias operacionais;
- a Secretaria de Planejamento e Investimentos Estratégicos/MPOG, em cumprimento às normas legais que impõem a avaliação anual dos programas do Plano Plurianual;
- a Secretaria de Avaliação e Gestão da Informação/MDS, em atenção à política de monitoramento e avaliação daquele Ministério;
- o IPEA, em cumprimento as suas competências legais.

Tais ações referem-se a avaliações externas, executadas por avaliadores que não integram as organizações responsáveis pela implementação dos programas, mesmo no caso do MDS, cujas avaliações são realizadas por entidades contratadas para esse fim.

Desta forma, considerando o tempo disponível para a realização desta pesquisa, que não permitiria aprofundar o estudo de todas as avaliações identificadas, optou-se por selecionar, inicialmente, as avaliações realizadas pela SPI/MPOG, SEPROG/TCU, SFC/CGU e SAGI/MDS, considerando que já haviam informações disponíveis, nos próprios sítios eletrônicos, sobre as sistemáticas adotadas para realização desses trabalhos, favorecendo a solicitação dos dados e informações necessários ao desenvolvimento desta pesquisa.

Em atenção às solicitações formuladas a esses órgãos foram disponibilizados os dados primários das avaliações de programas realizadas por intermédio dos questionários aplicados pela SPI nos exercícios de 2004 a 2007, além dos relatórios de avaliação, manuais de avaliação e demais orientações que se encontram disponíveis no sítio eletrônico do Ministério do Planejamento Orçamento e Gestão. Também foram disponibilizadas as avaliações e toda a documentação disponível em meio magnético relativa aos trabalhos realizados pela SEPROG/TCU, no mesmo período (2004-2007), bem como os manuais e demais documentos técnicos que dispõem sobre a estratégia metodológica utilizada para a realização dos estudos avaliativos.

Quanto aos dados e informações das avaliações realizadas pela SFC/CGU e pela SAGI/MDS, em que pese o interesse desses órgãos em colaborar com este trabalho, os dados e a documentação relativos às avaliações por eles realizadas não foram apresentados em tempo para análise, considerando, ainda, a limitação do prazo para conclusão desta pesquisa, que inviabilizou a reiteração das solicitações formuladas.

Desta forma, os casos foram selecionados entre as avaliações de programas realizadas pelo Tribunal de Contas da União e as avaliações dos programas do PPA 2004-2007 realizadas pelo Ministério do Planejamento Orçamento e Gestão.

Assim, as avaliações de programas foram definidas como unidade de análise para realização dos estudos de casos múltiplos e os padrões de avaliação como subunidades de análise a serem testadas com vistas à replicação dos casos.

Sobreleva mencionar, no entanto, que a lógica da análise dos casos múltiplos será realizada considerando, apenas, a replicação dos casos por órgão (PPA/SPI e TCU). Não será realizada análise transcaso das avaliações selecionadas para estudo, em razão da limitação de prazo para conclusão dos trabalhos.

3.3.1 – Da Seleção das Avaliações do TCU

No período 2004-2007, foram realizados 36 trabalhos de auditoria operacional, na modalidade avaliação de programas, coordenadas ou com a participação da Secretaria de Fiscalização e Avaliação de Programas do TCU (QUADRO 29).

Inicialmente, foram intencionalmente selecionadas 12 avaliações dentre as 36 realizadas (QUADRO 18), de forma a possibilitar a consulta ao maior número de avaliações conduzidas por diferentes coordenadores, auditores seniores, no período examinado, o que favoreceria a análise da qualidade e credibilidade das avaliações do TCU conduzidas por diferentes profissionais desse órgão.

Os coordenadores das auditorias são responsáveis pela condução do processo de avaliação, orientando, com o apoio do supervisor, todas as atividades desenvolvidas ao longo do planejamento, execução e elaboração dos relatórios das auditorias operacionais.

Exercício	Programas Selecionados
2004	Ação Assistência Financeira para Aquisição e Distribuição de Medicamentos Excepcionais
	Programa Combate ao Abuso e Exploração Sexual de Crianças e Adolescentes
2005	Programa Segurança Fitozoossanitária de Produtos Agropecuários
	Programa Segundo Tempo
	Ações de Fiscalização e de Acompanhamento de Ingressos de Terras Indígenas
	Programa Nacional de Erradicação da Hanseníase
	Programa Doação Captação e Transplante de Órgãos e Tecidos
	Programa Manutenção de Hidrovias
	Programa Governo Eletrônico
2006	Programa Desenvolvimento Sustentável da Aquicultura
	Ação Levantamentos Hidrogeológicos do Programa Geologia do Brasil
2007	Auditoria Operacional na Funasa (Coordenada pela Seprog em parceria)

QUADRO 18 - AVALIAÇÕES DE PROGRAMAS SELECIONADAS – TCU

Para definir os casos a serem estudados, foi encaminhado, em meio eletrônico, para os 12 coordenadores, convite para que participassem desta pesquisa, respondendo ao questionário elaborado para avaliação da observância dos padrões de avaliação definidos pelo *Joint Committee on Standards for Educational Evaluation*, constante do Apêndice. No entanto, só foram obtidas respostas ao convite para 7 avaliações. Desta forma, os casos selecionados do TCU para o desenvolvimento deste estudo foram as 7 avaliações relacionadas no QUADRO 19.

Avaliações Pesquisadas	Respondentes
Ação Assistência Financeira para Aquisição e Distribuição de Medicamentos Excepcionais	Coordenador e Supervisor
Programa Combate ao Abuso e Exploração Sexual de Crianças e Adolescentes	Coordenador
Ações de Fiscalização e de Acompanhamento de Ingressos em Terras Indígenas	Coordenador
Programa Doação Captação e Transplante de Órgãos e Tecidos	Coordenador
Programa Manutenção de Hidrovias	Coordenador
Programa Desenvolvimento Sustentável da Aquicultura	Coordenador e Supervisor
Ação Levantamentos Hidrogeológicos do Programa Geologia do Brasil	Coordenador

QUADRO 19 - RELAÇÃO DAS AVALIAÇÕES EXAMINADAS E RESPECTIVOS RESPONDENTES

3.3.2 – Da Seleção das Avaliações de Programas do PPA/SPI/MPOG

No que diz respeito às avaliações dos programas do PPA 2004-2007, preliminarmente, procedeu-se a análise das avaliações dos programas constantes do Plano, mediante a identificação e quantificação dos programas avaliados no período selecionado para exame, tendo sido identificados 419 programas. Desse total, foram selecionados os programas avaliados durante os quatro exercícios, tendo-se identificado 232 programas. Desse conjunto,

selecionou-se uma amostra não estatística de programas, cujo exame das avaliações permitisse verificar, por intermédio da evolução do desempenho e dos resultados, as melhorias e aperfeiçoamentos que poderiam indicar a utilização das avaliações para o alcance dos objetivos originalmente estabelecidos. Adotou-se como critério para escolha desses programas, o programa de maior materialidade por órgão, desde que houvesse mais de um programa sob a responsabilidade de um mesmo órgão. Como resultado, foram selecionados os 26 programas relacionados no QUADRO 20.

Órgão Responsável	Programas PPA 2004-2007 Selecionados
PR	Democratização do Acesso à Informação Jornalística, Educacional e Cultural
SEDH	Direitos Humanos, Direitos de Todos
SPM	Prevenção/Combate à Violência Contra as Mulheres
MAPA	Abastecimento Agroalimentar
MCT	Nacional de Atividades Nucleares
MF	Desenvolvimento do Sistema Financeiro Nacional
MEC	Universidade do Século XXI
MDIC	Desenvolvimento do Comércio Exterior/ e da cultura Exportadora
MJ	Combate à Criminalidade
MME	Atuação Internacional na Área de Petróleo
MPS	Previdência Social Básica
MRE	Relações do Brasil com Estados Estrangeiros
MS	Atenção Hospitalar e Ambulatorial no Sistema Único de Saúde
TEM	Integração das Políticas Públicas de Emprego, Trabalho e Renda
MT	Manutenção da Malha Rodoviária Federal
MC	Aprimoramento dos Serviços Postais
MinC	Engenho das Artes
MMA	Recursos Pesqueiros Sustentáveis
MPOG	Recenseamentos Gerais
MDA	Agricultura Familiar - PRONAF
ME	Esporte e Lazer na Cidade
MD	Serviço de Saúde das Forças Armadas
MIN	Pro-água Infraestrutura
MTur	Turismo no Brasil: Uma Viagem para Todos
MDS	Transferência de Renda com Condiionalidades
Mcidades	Descentralização dos Sistemas de Transporte Ferroviário Urbano de Passageiros

QUADRO 20 - PROGRAMAS SELECIONADOS DO PPA 2004-2007 PARA EXAME DAS AVALIAÇÕES

3.4 ANÁLISE DOS DADOS

Para o exame das questões desta pesquisa, definiu-se como subunidades de análise, os padrões de avaliação estabelecidos pelo Comitê Conjunto de Padrões para Avaliação Educacional, com o objetivo de confrontá-los com os casos (as avaliações de programas selecionadas acima) e, desta forma, verificar em que medida as sistemáticas e abordagens metodológicas adotadas, para a realização das avaliações de programas examinadas, são

consistentes com a prática avaliativa, com qualidade e credibilidade, favorecendo, desta forma, a utilização dos resultados, uma vez que os padrões estão relacionados aos fatores que influenciam a utilização, conforme apresentado no QUADRO 17.

Para tanto, foi elaborado questionário, constante do Apêndice, com base em Sanders (1994), para avaliação do nível de observância dos padrões, o qual foi encaminhado por meio eletrônico para os 12 coordenadores e 3 supervisores das avaliações selecionadas do TCU e mais 2 especialistas, também do Tribunal, bem como para o responsável pela área de Coordenação da Qualidade do Plano da SPI/MPOG, solicitando-lhes que atribuíssem nota de 1 a 4 para cada questão, de acordo com a percepção e a documentação relativa aos trabalhos por eles coordenados e supervisionados, observando-se as especificações para atribuição das notas conforme disposto no QUADRO 21.

Escala de Notas	Especificação
NA	Não aplicável
1	Não Observado
2	Observado com restrições que limitam o pleno atendimento da questão examinada
3	Observado com restrições, mas atende à questão examinada
4	Observado

QUADRO 21 - ESCALA DE AVALIAÇÃO DE OBSERVÂNCIA DOS PADRÕES

Cabe observar, no entanto, que foram recebidas respostas para o questionário apenas de 7 coordenadores e 2 supervisores, no caso do TCU e do responsável pela Coordenação da Qualidade do Plano da SPI/MPOG. Desta forma, tem-se que para duas, das sete avaliações selecionadas, realizadas pelo TCU - Programa Desenvolvimento Sustentável da Aquicultura e Ação Assistência Financeira para Aquisição e Distribuição de Medicamentos Excepcionais - foram recebidas respostas tanto dos coordenadores como dos supervisores das avaliações, totalizando 9 respondentes.

Precedeu o envio do questionário, com vistas a apresentar os objetivos desta pesquisa e os instrumentos de coleta de dados, a realização de reuniões com 2 coordenadores e 2 supervisores dos trabalhos relacionados, além de entrevistas com o titular e assessora da Secretaria de Fiscalização e Avaliação de Programas/TCU, bem como com o Secretário e o Diretor responsável pela área de Coordenação da Qualidade do Plano da SPI/MPOG.

Para o exame das avaliações de programas do PPA 2004/2007, com vistas a caracterizar a tipologia de avaliação e a adequação da sistemática e da metodologia adotadas aos objetivos propostos, procedeu à análise dos relatórios de avaliação dos 26 programas relacionados no QUADRO 20, relativos aos quatro exercícios, em confronto com os manuais e normas que regulamentaram a sua implementação. Além disso, os relatórios de avaliação

foram ainda confrontados como as respostas constantes da BASE DE DADOS AVALIAÇÃO_PPA2004-2007 (SPI/MPOG), com o objetivo de evidenciar a evolução das respostas apresentadas às questões, formuladas na avaliação anual dos programas, referentes à concepção, à implementação e aos resultados.

O resultado dessa análise foi comparado às respostas do responsável pela área de Coordenação da Qualidade do Plano da SPI/MPOG, ao questionário aplicado nesta pesquisa.

Com relação às avaliações do TCU, procedeu-se à análise dos 7 relatórios de avaliação selecionados em confronto com as respostas encaminhadas ao questionário (APÊNDICE) por 7 coordenadores e 2 supervisores, como também com manuais e documentos técnicos relativos à sistemática e abordagem metodologia aplicadas na realização dos trabalhos, relacionados no QUADRO 22.

Relação de Documentos Técnicos do TCU Utilizados na Pesquisa
Manual de Auditoria Operacional (2000)
Roteiro para Elaboração de Relatórios de Auditoria Operacional
Controle de Qualidade de Auditorias de Natureza Operacional
Indicadores de Desempenho e Mapa de Produtos
Mapa de Processos
Análise RECI
Marco Lógico
Análise Stakeholder
Análise SWOT e Matriz de Verificação de Risco
Técnica de Entrevista
Técnica de Apresentação de Dados

QUADRO 22 - DOCUMENTOS TÉCNICOS DO TCU RELATIVOS À SISTEMÁTICA E ABORDAGEM METODOLÓGICA APLICADAS NAS AUDITORIAS DE NATUREZA OPERACIONAL

Adicionalmente a esse procedimento e com vistas a obter elementos para avaliar as respostas recebidas, além dos relatórios das avaliações relacionadas no QUADRO19, foi examinada a documentação disponível em meio magnético relativa a três avaliações, compilada nas fases de seleção, planejamento, execução e relatório dos trabalhos referentes à Ação Assistência Financeira para Aquisição e Distribuição de Medicamentos Excepcionais; ao Programa Doação Captação e Transplante de Órgãos e Tecidos e ao Programa Manutenção de Hidrovias (QUADRO 23). Assim, aplicou-se a essas avaliações o questionário constante do APÊNDICE, tendo sido atribuídas notas por esta pesquisadora, segundo a escala de notas constante do QUADRO 21.

Avaliação	Documentos do TCU Revisados por Fase do Ciclo de Trabalho das Avaliações			
	Seleção	Planejamento	Execução	Elaboração do Relatório
Medicamentos	Relatórios de Caracterização e Estudo de Viabilidade	<i>Técnicas de Diagnóstico</i>	Extração Respostas Pesquisa_Análise Estatística	Relatório Final antes dos comentários do gestor
		Mapa de Produto	Matriz de Achados	Relatório Final
		Mapa de Processo		
		Marco Lógico Alto Custo		
		Marco Lógico Assistência Farmacêutica		
		Marco Lógico AFB e PFMFI		
		Análise RECI		
		Análise Stakeholder		
		<i>Matriz de Planejamento</i>		
		<i>Instrumentos de Coleta de Dados</i>		
		Questionários Associação de Beneficiários		
		Questionários Beneficiários		
		Questionários Laboratórios Oficiais		
		Roteiros de Entrevista Gestor Estadual		
		Roteiros de Entrevista CRM e Associações		
		Roteiros de Entrevista Dispensadores		
		Roteiros de Visita		
<i>Relatório de Planejamento</i>				
Transplante	Relatórios de Caracterização e Estudo de Viabilidade	<i>Técnicas de Diagnóstico</i>		Relatório Final antes dos comentários do gestor
		Análise de Indicadores de Desempenho		Relatório Final
		Análise Stakeholder		
		Análise do Orçamento		
		Diagrama de Verificação de Risco		
		Mapa de Processo	Matriz de Achados	
		Mapa de Produto		
		Matriz Coso		
		Análise SWOT		
		<i>Matriz de Planejamento</i>		
		<i>Metodologia de Pesquisa Aplicada</i>		
		<i>Relatório de Planejamento</i>		

(continua)

(conclusão)

Avaliação	Documentos do TCU Revisados por Fase do Ciclo de Trabalho das Avaliações				
	Seleção	Planejamento	Execução	Elaboração do Relatório	
Hidroviás	Relatórios de Caracterização e Estudo de Viabilidade	<i>Técnicas de Diagnóstico</i>	Respostas aos questionários	Relatório Final antes dos comentários do gestor	
		Análise Stakeholder	Matriz de Achados	Relatório Final	
		Análise SWOT			
		Avaliação dos Indicadores de Desempenho			
		Diagrama de Verificação de Risco			
		Mapa de Processo			
		Mapa de Produto			
		<i>Instrumentos de Coleta de Dados</i>			
		Questionário ONG			
		Questionário Postal Usuários			
		Roteiro de Entrevista Capitania dos Portos			
		Roteiro de Entrevista Ministério dos Transportes			
		Roteiro de Entrevista Administrações Hidroviárias			
		Roteiro de Entrevista ANTAq			
		Roteiro de Entrevista Docas			
		Roteiro de Entrevista IBAMA			
		Roteiro de Entrevista ONGs			
		Roteiro de Entrevista Transportadoras			
		Roteiro de Informações requisitadas da Marinha			
		Roteiro de Observação Direta Hidroviás			
		Roteiros de Viagem			
Registros de entrevistas e reuniões					
<i>Matriz de Planejamento</i>					
<i>Relatório de Planejamento</i>					

QUADRO 23 - DOCUMENTAÇÃO DAS AVALIAÇÕES DO TCU EXAMINADA POR ESTA PESQUISADORA

Para as avaliações do TCU, foi possível calcular a média das notas atribuídas pelos coordenadores e supervisores às questões formuladas no questionário para cada padrão (APÊNDICE), bem como calcular a média das notas atribuídas por esta pesquisadora no exame da documentação relativa às 3 avaliações mencionadas anteriormente e a média geral dessas notas.

Com base nas médias das notas, foram considerados adequados aqueles padrões cuja média geral situou-se entre 3 e 4, que indicam que o padrão foi observado, mesmo com alguma restrição, mas que não comprometeu o resultado desejado.

Para o cálculo das médias não foram consideradas as respostas avaliadas como NA (não aplicável) e aquelas não preenchidas pelos respondentes.

No QUADRO 24, estão relacionados os 30 padrões examinados nesta pesquisa por intermédio das questões relacionadas no Questionário para Avaliação da Observância dos

Padrões de Avaliação Definidos pelo *Joint Committee on Standards for Educational Evaluation*, constante do Apêndice.

Padrões	DENOMINAÇÃO
Utilidade	
U1	IDENTIFICAÇÃO STAKEHOLDER
U2	CREDIBILIDADE AVALIADOR
U3	IDENTIFICAÇÃO DAS INFORMAÇÕES NECESSÁRIAS PARA SELEÇÃO E DEFINIÇÃO DO ESCOPO
U4	IDENTIFICAÇÃO DOS VALORES (CRITÉRIOS)
U5	CLAREZA DO RELATÓRIO
U6	TEMPESTIVIDADE E DISSEMINAÇÃO DO RELATÓRIO
U7	IMPACTO DA AVALIAÇÃO (INFLUÊNCIA PARA TOMADA DE DECISÃO)
Viabilidade	
F1	PROCEDIMENTOS PRÁTICOS
F2	VIABILIDADE POLÍTICA
F3	CUSTO-EFETIVIDADE
Propriedade	
P1	ORIENTAÇÃO PARA PRESTAÇÃO DOS SERVIÇOS
P2	ACORDO FORMAL
P3	QUESTÕES RELATIVAS A DIREITOS INDIVIDUAIS
P4	RELAÇÕES HUMANAS
P5	AVALIAÇÃO COMPLETA E IMPARCIAL (JUSTA)
P6	APRESENTAÇÃO DOS ACHADOS
P7	CONFLITO DE INTERESSES
P8	RESPONSABILIDADE FISCAL
Acuracidade	
A1	DOCUMENTAÇÃO DO PROGRAMA
A2	ANÁLISE DO CONTEXTO
A3	DESCRIÇÃO DOS PROPÓSITOS E PROCEDIMENTOS
A4	FONTES DE INFORMAÇÕES DEFENSÁVEIS
A5	INFORMAÇÕES VÁLIDAS
A6	INFORMAÇÕES CONFIÁVEIS
A7	INFORMAÇÕES SISTEMATIZADAS
A8	ANÁLISE QUANTITATIVA DAS INFORMAÇÕES
A9	ANÁLISE QUALITATIVA DAS INFORMAÇÕES
A10	CONCLUSÕES JUSTIFICADAS
A11	RELATÓRIO IMPARCIAL
A12	META-AVALIAÇÃO

QUADRO 24 - PADRÕES DE AVALIAÇÃO DE PROGRAMAS DEFINIDOS PELO COMITÊ CONJUNTO DE PADRÕES PARA AVALIAÇÃO EDUCACIONAL DOS EUA

3.5 COLETA DE DADOS

Para coleta dos dados elaborou-se o questionário constante do Apêndice, segundo Sanders (1994), no qual para cada Padrão relacionado no QUADRO 24 foram definidas

algumas questões, que objetivam identificar a observância do respectivo padrão. Esse questionário foi encaminhado para os coordenadores das avaliações do TCU, mencionados anteriormente, e para o responsável pela área de Coordenação da Qualidade do Plano da SPI/MPOG, solicitando-lhes que atribuíssem nota de 1 a 4 para cada questão, de acordo com a sua percepção e baseados na documentação relativa aos trabalhos por eles coordenados e supervisionados. Assim, como exemplo, apresenta-se no QUADRO 25, as questões formuladas para os Padrões de Utilidade (U1) – Identificação dos Stakeholders e (U2) – Credibilidade do Avaliador. O questionário com todas as questões examinadas consta do Apêndice - Questionário – Avaliação da Observância dos Padrões de Avaliação Definidos pelo *Joint Committee On Standards For Educational Evaluation*

Orientações	Questões	Nota	Comentário
Pessoas envolvidas na avaliação ou por ela afetadas devem ser identificadas para que suas necessidades possam ser contempladas.	U1 - Identificação Stakeholder		
	Os destinatários da avaliação foram identificados?		
	Os stakeholders relevantes para definição do problema foram identificados?		
	Existe um processo (rotina) para identificação dos stakeholders?		
	Existe a definição das atribuições de cada stakeholder?		
	Os stakeholders identificados participaram do processo de desenho da avaliação?		
	Os stakeholders participaram do processo de execução da avaliação?		
	As demandas de informações dos principais stakeholders foram atendidas?		
Pessoas conduzindo avaliação devem ser competentes e confiáveis para a sua realização.	U2 - Credibilidade do Avaliador		
	Existem critérios para selecionar os avaliadores?		
	A equipe de avaliadores e o plano de trabalho foram avaliados?		
	O Plano da avaliação foi apresentado para os diversos stakeholders e foi considerando realístico e tecnicamente adequado para responder às demandas apresentadas pelos stakeholders?		
	Foram identificados os atores-chave para acompanhar o progresso dos trabalhos de avaliação?		

QUADRO 25 - EXTRATO DO QUESTIONÁRIO APLICADO NA PESQUISA PARA COLETA DE DADOS SOBRE A OBSERVÂNCIA DOS PADRÕES DE AVALIAÇÃO DE PROGRAMAS DEFINIDOS PELO COMITÊ CONJUNTO DE PADRÕES PARA AVALIAÇÃO EDUCACIONAL DOS EUA

Com base na análise das respostas obtidas com a aplicação do questionário (APÊNDICE), bem como das informações obtidas nas entrevistas e no levantamento documental foi possível evidenciar se as avaliações governamentais realizadas pela

SPI/MPOG e pelo TCU favorecem ou não a utilização dos resultados, conforme descrito nos capítulos 5 e 6.

3.6 DAS LIMITAÇÕES DA ESTRATÉGIA METODOLÓGICA

As estratégias metodológicas aplicadas para o estudo da utilização da avaliação têm merecido ressalvas desde o trabalho realizado por Leviton e Hughes (1981), onde um dos pontos abordados pelos autores foi, justamente, a questão dos métodos empregados no estudo da utilização, cujas principais abordagens adotadas são os estudos de caso e as *surveys*. Esses métodos, segundo os autores, enfrentam pelo menos quatro problemas: 1) a dificuldade em documentar que a utilização ocorreu, haja vista que as avaliações são frequentemente utilizadas informalmente e o tempo decorrido entre a utilização e a realização do estudo pode comprometer a recuperação das informações necessárias para o exame; 2) a dificuldade em demonstrar que a mudança, em qualquer nível, foi causada pelo menos em parte pela avaliação; 3) a questão das bases para comparação; e, por último, 4) as deficiências na identificação da unidade de análise.

Na pesquisa realizada por Cousins e Leithwood (1986), também foram examinadas as características metodológicas dos estudos empíricos para investigar a utilização, sendo que os 65 trabalhos por eles analisados, utilizaram os estudos retrospectivo e longitudinal e a pesquisa simulativa.

Os estudos retrospectivos focaram em avaliações previamente realizadas e se basearam principalmente em dados obtidos das memórias dos decisores, mantenedores e/ou avaliadores, sendo a sua maioria coletados por intermédio de *surveys* e alguns incluíram estudos de casos únicos e múltiplos. Já os estudos longitudinais examinaram a influência dos dados obtidos antes, durante ou depois da realização da avaliação, utilizando os métodos quase-experimentais. E, nas pesquisas simulativas, foram utilizados, em geral, os métodos experimentais para testar os efeitos antecipados da utilização relativos a relatórios de avaliação específicos, sob condições altamente controladas (Cousins e Leithwood, 1986).

No entanto, segundo Brandon e Singh (2009), ao examinarem os métodos empregados nos estudos empíricos sobre a utilização dos achados das avaliações de programas, a partir dos estudos mais referenciados na literatura, ou seja, aqueles conduzidos por Leviton e Hughes (1981), Thompson e King (1981), Cousins e Leithwood (1986), Shulha e Cousins (1997) e Cousins (2003), foi apontado que, com exceção do artigo de Cousins e

Letihwood, todos os demais são revisões tradicionais da literatura, que não incluem análise quantitativa dos resultados dos estudos.

Na pesquisa realizada por esses autores sobre os métodos aplicados nos estudos empíricos sobre a utilização das avaliações, eles concluem que “quando os trabalhos são considerados conjuntamente com outros achados sobre o conhecimento da utilização em geral, e são encontradas similaridades entre os conjuntos de achados, as deficiências metodológicas são menos importantes; mas quando consideradas apenas como resultado da utilização da avaliação, os achados dos estudos por eles examinados, como um todo, não tem suficiente credibilidade científica (BRANDON e SINGH, 2009, p. 135, tradução nossa).

Essas evidências demonstram as deficiências metodológicas do estudo da utilização e reforçam a importância da pesquisa bibliográfica para assegurar a credibilidade da pesquisa empírica. Assim, como no Brasil os estudos sobre a utilização das avaliações ainda não se encontram sistematizados de forma a permitir a sua recuperação para realização de uma pesquisa nos moldes daquelas já realizadas em outros países, neste trabalho, procurou-se delinear uma proposta que favoreça, primeiramente, o exame das questões conceituais para propiciar, posteriormente, a investigação da utilização propriamente dita.

4 A AVALIAÇÃO DE PROGRAMAS GOVERNAMENTAIS NO BRASIL

No âmbito da administração pública brasileira, a função avaliação ganhou destaque na Constituição Federal de 1988, na seção que trata da fiscalização contábil, financeira e orçamentária, como é competência expressa do sistema de controle interno dos poderes constituídos, tendo, dentre suas finalidades, a de avaliar o cumprimento das metas previstas no plano plurianual, os orçamentos da União, bem como a execução dos programas de governo e a eficácia e eficiência de seus resultados e da gestão dos órgãos e entidades governamentais (BRASIL, 1988).

No entanto, a função avaliação foi impulsionada com a introdução do modelo de administração pública gerencial, proposto no Plano Diretor da Reforma do Aparelho do Estado de 1995, cujo paradigma fundamenta-se no controle dos resultados, objetivando a melhoria da eficiência e da qualidade na prestação dos serviços públicos, bem como no desenvolvimento de uma cultura gerencial nas organizações (BRASIL, 1995).

A exposição de motivos 49/95, que acompanhou a proposta de emenda constitucional que trata da reforma administrativa, encaminhada ao Congresso Nacional em 23 de agosto de 1997, destacou esses pressupostos, nos seguintes termos:

No difícil contexto do retorno à democracia, que em nosso país foi simultâneo à crise financeira do Estado, a Constituição de 1988 corporificou uma concepção de administração pública verticalizada, hierárquica, rígida, que favoreceu a proliferação de controles muitas vezes desnecessários. Cumpre agora, reavaliar algumas das opções e modelos adotados assimilando novos conceitos que reorientem a ação estatal em direção à eficiência e à qualidade dos serviços prestados ao cidadão.

A revisão de dispositivos constitucionais e inúmeras outras mudanças na esfera jurídico-legal que a acompanharão, estão direcionadas para o delineamento de condições propícias à implantação de novos formatos organizacionais e institucionais, à revisão de rotinas e procedimentos e à substituição dos controles formais pela avaliação permanente de resultados (BRASIL, 1997).

Desta forma, em 1996, foi lançado o Programa Brasil em Ação, uma proposta do governo que agrupou 42 empreendimentos, tanto da área social, como de infraestrutura, voltados para reduzir custos e aumentar a competitividade do setor produtivo, cujo modelo de gerenciamento se propunha a assegurar a tempestividade na execução das ações, com base na solução antecipada de problemas e no controle dos resultados (BRASIL, 1998a).

Posteriormente, o modelo de planejamento e gestão das ações governamentais, embrionariamente desenvolvido no Brasil em Ação, foi adotado na elaboração do Plano Plurianual 2000-2003 e nos demais orçamentos da União, a partir do exercício de 2000, tendo as diretrizes para sua elaboração sido estabelecidas no Decreto n. 2.829/98 (BRASIL, 1998b).

O mencionado Decreto determinou a obrigatoriedade de avaliação anual da consecução dos objetivos estratégicos do Governo Federal e dos resultados dos programas, com a finalidade de aferir o cumprimento dos objetivos e das metas fixadas, subsidiar o processo de alocação de recursos públicos, a política de gastos públicos e a coordenação das ações de governo, bem como evitar a dispersão e o desperdício de recursos públicos (BRASIL, 1998b).

No que diz respeito à gestão da qualidade, as unidades responsáveis pela execução dos programas deveriam manter, quando cabível, sistema de avaliação do grau de satisfação da sociedade quanto aos bens e serviços ofertados pelo Poder Público (BRASIL, 1998b).

Além desses normativos, a Lei de Responsabilidade Fiscal estabeleceu ainda que a avaliação dos resultados dos programas do PPA financiados com recursos orçamentários deveria observar, anualmente, as disposições da Lei de Diretrizes Orçamentárias.

Desta forma, a partir da elaboração do PPA 2000-2003 (Avança Brasil), buscou-se integrar as funções planejamento, orçamento, gestão e avaliação, tendo os programas de governo como elemento central de estruturação das ações finalísticas do governo.

As mudanças introduzidas no âmbito político, econômico, administrativo e institucional no Brasil, nos anos 80-90, impuseram, de igual modo, transformações nas organizações públicas responsáveis pelo cumprimento das novas demandas impostas pelo Estado, tanto nas suas estruturas, como nos seus processos de trabalho.

Nesse sentido, a Medida Provisória n. 480, de 27.04.1994, posteriormente convertida na Lei n. 10.180, de 06.02.2001, organizou os sistemas de planejamento e de orçamento, de administração financeira e de contabilidade federal, bem como de controle interno do Poder Executivo Federal. Assim, coube ao Ministério do Planejamento, Orçamento e Gestão, como também às demais unidades responsáveis pelas atividades de planejamento, a competência, dentre outras, de avaliar os planos e programas de governo quanto à eficácia e efetividade, com vistas a subsidiar o processo de alocação de recursos públicos, a política de gastos e a coordenação. Além disso, também assegurar que as unidades responsáveis pela execução dos programas, projetos e atividades da administração federal mantivessem rotinas de acompanhamento e avaliação da sua programação (BRASIL, 2001a).

Além dessas entidades, coube ao Instituto de Pesquisas Econômicas Aplicadas - IPEA, vinculado à Secretaria de Assuntos Estratégicos da Presidência da República, conforme estabelecido no Decreto 4.745/2003 – a competência de realizar pesquisas e estudos sociais e econômicos, disseminar o conhecimento resultante, dar apoio técnico e institucional ao Governo na avaliação, formulação e acompanhamento de políticas públicas, planos e

programas de desenvolvimento, bem como oferecer à sociedade elementos para o conhecimento e solução dos problemas e dos desafios do desenvolvimento brasileiro (BRASIL, 2003).

Desta forma, o IPEA realiza as seguintes atividades:

- i) pesquisas destinadas ao conhecimento dos processos econômicos e sociais brasileiros;
- ii) análise e diagnóstico dos problemas estruturais e conjunturais da economia e da sociedade brasileira;
- iii) estudos prospectivos de médio e longo prazo;
- iv) fornecimento de subsídios técnicos para a formulação de políticas públicas e para a preparação de planos e programas de governo;
- v) análise e avaliação de políticas públicas, programas e ações governamentais;
- vi) capacitação técnica e institucional para o aperfeiçoamento das atividades de planejamento, avaliação e gestão; e
- vii) disponibilização de sistemas de informação e disseminação de conhecimentos atinentes às suas áreas de competência.

Ao sistema de controle interno do Poder Executivo, conforme disposto no Art. 19. da Lei n. 10.180/2001, coube a avaliação da ação governamental e da gestão dos administradores públicos federais, por intermédio da fiscalização contábil, financeira, orçamentária, operacional e patrimonial (Brasil, 2001a). A regulamentação do sistema de controle interno do poder executivo foi estabelecida pelo Decreto n. 3.591 de 6 de setembro de 2000.

Cabe observar, no entanto, que o Brasil adota o sistema duplo de controle da administração pública Federal: o controle externo, de competência do Congresso Nacional, com o auxílio do Tribunal de Contas da União; e o controle interno, exercido pelo sistema de controle interno de cada Poder (BRASIL, 1988).

Assim, compete ao Tribunal de Contas da União - TCU, no exercício do controle externo – normatizar e realizar inspeções e auditorias de natureza contábil, financeira, orçamentária, operacional e patrimonial, com vistas a verificar a legalidade, a economicidade, a legitimidade, a eficiência, a eficácia e a efetividade de atos, contratos e fatos administrativos praticados no âmbito da administração pública federal. (BRASIL, 1992 e 2002).

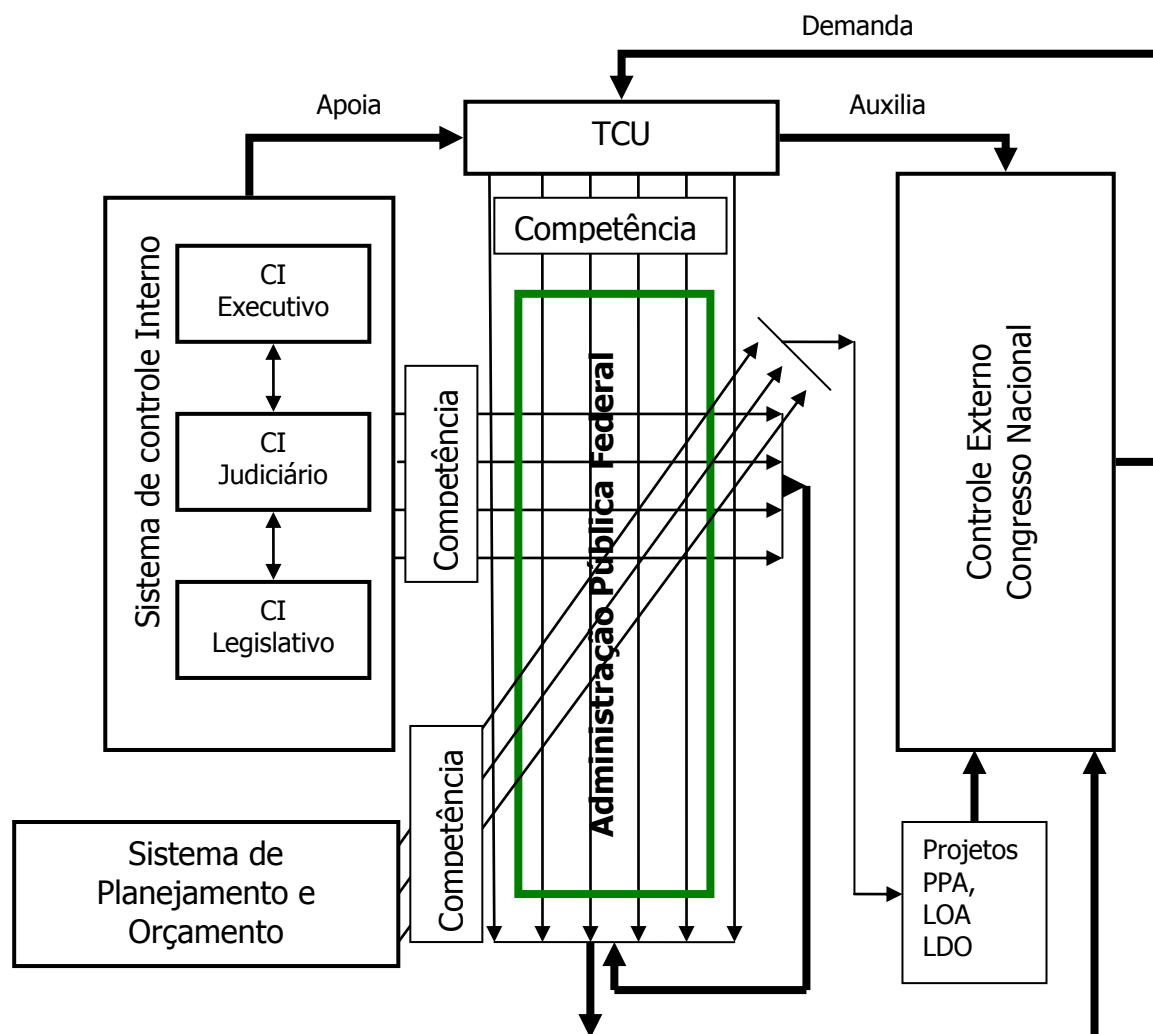


FIGURA 6 - ATUAÇÃO DOS SISTEMAS DE CONTROLE INTERNO, EXTERNO E DO SISTEMA DE PLANEJAMENTO E ORÇAMENTO

Pode-se depreender, dos dispositivos legais mencionados, que a função avaliação foi instituída no país tanto como instrumento para subsidiar e retroalimentar as funções de planejamento, orçamento e gestão governamentais, como também para o controle da administração pública, que se realiza por meio das fiscalizações e auditorias, tanto no âmbito do sistema de controle interno, como do controle externo, a cargo do TCU (FIGURA 6).

Esta constatação mostra-se coerente com o observado em outros países, conforme nos informa Rist (Ed., 1990), que, ao examinar o desenvolvimento, institucionalização e utilização da avaliação em oito países democráticos (Canadá, Alemanha, Reino Unido, Estados Unidos, Dinamarca, Holanda, Noruega e Suíça), constatou a vinculação da função avaliação com as competências dos órgãos governamentais de auditoria.

Segundo Rist (Ed., 1990, p. 8-9) pode-se depreender que se trata de um casamento forçado, considerando que auditoria e avaliação não são a mesma coisa, tampouco têm a

mesma origem. No entanto, o autor apresenta dois fatores que podem justificar essa constatação. O primeiro diz respeito a demandas de informações dos formuladores e executores de políticas e programas, que, face às restrições fiscais, já não se contentavam apenas em saber se os recursos foram corretamente aplicados (conformidade legal), passando a requerer informações sobre a implementação (desempenho) e os resultados. O segundo fator refere-se à questão metodológica, considerando que a avaliação tem caminhado para o desenvolvimento de métodos normativos e a auditoria tem ampliado o escopo de seus trabalhos buscando a análise causal; desta forma as duas funções acabam por compartilhar interesses comuns, favorecendo a interação entre os dois instrumentos.

As reformas gerenciais do Estado das últimas décadas favoreceram a ampliação das competências dos órgãos de auditoria governamentais, conforme observado por Barzelay (1997), introduzindo novas modalidades de auditoria, dentre elas a auditoria de desempenho (*performance auditing*), que rapidamente foi aceita como um tipo de revisão da atividade governamental *a posteriori*.

O General Accountability Office (GAO), nos Estados Unidos, em 1981, por exemplo, na primeira revisão dos padrões de auditoria de organizações governamentais, programas, atividades e funções (o livro amarelo), apresentou, como uma das justificativas para a revisão, a necessidade de separar os padrões de auditoria financeira e legal daqueles de auditoria de economia e eficiência e de auditoria de resultado de programas, justificando tal medida nos seguintes termos:

Nos anos 60, os programas de governo ou os recursos financeiros a eles destinados cresceram substancialmente. Tais programas destinavam-se a melhorar a qualidade de vida americana. Nos anos 70, o número de novos programas estabelecidos foi consideravelmente menor, mas os recursos financeiros continuaram a crescer. Tal crescimento nos programas de governo trouxe com ele um aumento da exigência de plena responsabilidade contábil dos encarregados de administrar os programas.

A auditoria é parte integral de tal responsabilidade contábil e os governos devem assegurar que se façam as auditorias apropriadas. Nos últimos anos, tem havido um interesse sem precedentes quanto à auditoria do governo. Os funcionários públicos, os legisladores e cidadãos comuns, querem e têm o direito de saber, não somente como verbas governamentais são manuseadas e se as leis e os regulamentos são cumpridos, como também se as organizações governamentais estão alcançando os objetivos pelos quais os programas foram criados e subvencionados e se estão se desempenhando com economicidade e eficiência (EUA, 1984).

A auditoria de desempenho surgiu como um instrumento de fiscalização voltado para a aferição dos resultados das ações governamentais. Seu principal objetivo, conforme definido por Barzelay (1997) ao proceder a análise comparativa sobre a auditoria de desempenho praticada pelas instituições de auditoria governamental dos países membros da OCDE, pode

ser expresso como *performance accountability*, que se relaciona à responsabilização das organizações pelos resultados de suas atividades.

Barzelay (1997) constatou ainda, que a auditoria de desempenho apresenta características similares à avaliação de programas, levando-o inclusive a concluir que essa modalidade de auditoria é, na realidade, avaliação de programa, porque envolve a produção e revisão de instrumental de julgamento.

Segundo os padrões definidos pela Internacional *Organization of Supreme Audit Institutions* - INTOSAI, a “auditoria de desempenho é o exame independente da eficiência e eficácia das ações, programas ou organizações governamentais, levando-se em consideração a economia, com vistas a promover o seu aperfeiçoamento.” (INTOSAI, 2004, p. 11, tradução nossa).

No Brasil, a modalidade de auditoria voltada à aferição do desempenho e dos resultados da ação governamental, quanto à economia, eficiência, eficácia e efetividade, denomina-se auditoria operacional, conforme expresso na Constituição Federal e definido nos normativos do Tribunal de Contas da União.

Com base na revisão normativa apresentada, em síntese, as avaliações governamentais podem ser classificadas de acordo com a seguinte tipologia:

Tipo de Avaliação	Objeto (o que é?)	Objetivo (para quê?)	Quem faz? Unidade Avaliadora	
Dec 2829/1998				
Art. 5				
Avaliação de objetivos estratégicos	Avaliar a consecução dos objetivos estratégicos do Governo Federal.	Subsidiar a elaboração da LDO.	MPOG e Unidades Responsáveis pelas Atividades de Planejamento	
Avaliação de resultados de programas	Avaliar os resultados dos programas.			
Art. 6				
Avaliação física e financeira de programas	Avaliar física e financeiramente os programas	Aferir o resultado quanto ao cumprimento dos objetivos e das metas fixadas.		
		Subsidiar o processo de alocação de recursos.		
		Subsidiar a política de gastos públicos.		
		Subsidiar a coordenação das ações de governo.		
		Evitar a dispersão e o desperdício de recursos públicos.		
Art. 7				
Avaliação da Qualidade dos Serviços Públicos	Gestão da qualidade	Satisfação da sociedade quanto aos bens e serviços ofertados pelo Poder Público.		

(continua)

(conclusão)

Tipo de Avaliação	Objeto (o que é?)	Objetivo (para quê?)	Quem faz? Unidade Avaliadora
Lei 10180/2001 Art. 7 e Decreto 4745/2003			
Acompanhamento físico e financeiro dos programas	Acompanhar física e financeiramente os planos e programas de governo.	Subsidiar o processo de alocação de recursos públicos; Subsidiar a política de gastos; Subsidiar a coordenação das ações do governo.	MPOG e Unidades Responsáveis pelas Atividades de Planejamento
Avaliação da eficácia e efetividade de programas	Avaliar planos e programas de governo quanto à eficácia e efetividade.		
Avaliação de investimentos	Avaliar investimentos estratégicos do governo.		
Estudos e pesquisas sócio-econômicas	Realizar estudos e pesquisas sócio-econômicas e análises de políticas públicas.	Dar apoio técnico e institucional ao Governo na avaliação, formulação e acompanhamento de políticas públicas, planos e programas de desenvolvimento e oferecer à sociedade elementos para o conhecimento e solução dos problemas e dos desafios do desenvolvimento brasileiro.	IPEA
CF, Lei 10180/2001 Art. 20 e Dec 3591/2000			
Avaliação de metas PPA	Avaliar o cumprimento das metas previstas no plano plurianual.	Comprovar a conformidade da sua execução.	Sistema de Controle Interno dos Poderes Legislativo, Executivo e Judiciário
Avaliação dos orçamentos	Avaliar os orçamentos da União.	Comprovar a conformidade da execução com os limites e destinações estabelecidos na legislação pertinente.	
Avaliação da execução dos programas	Avaliar a execução dos programas de governo.	Comprovar o nível de execução das metas, o alcance dos objetivos e a adequação do gerenciamento.	
Avaliação de eficácia e eficiência da gestão	Avaliar os resultados quanto à eficácia e eficiência da gestão orçamentária, financeira e patrimonial nos órgãos e entidades da Administração	Comprovar a legalidade e a legitimidade dos atos; examinar os resultados quanto à economicidade, à eficiência e à eficácia da gestão orçamentária, financeira, patrimonial, de pessoal e demais sistemas administrativos e operacionais.	
CF, Lei Orgânica TCU e RITCU			
Auditoria Operacional	Exame da economicidade, eficiência, eficácia e efetividade de organizações, programas e atividades governamentais.	Obter melhorias em organizações, programas e atividades governamentais.	TCU

QUADRO 26 - TIPOLOGIA DE AVALIAÇÃO GOVERNAMENTAL

Conforme se observa do QUADRO 26, as avaliações governamentais no Brasil têm como principal objetivo subsidiar o processo orçamentário, oferecendo contribuições para a melhoria da alocação, por meio das avaliações de metas físicas e financeiras. Essas avaliações são, em geral, consolidadas nas avaliações do plano plurianual e no relatório que acompanha o Balanço Geral da União.

Outro objetivo depreendido da tipologia apresentada no QUADRO 26, diz respeito à avaliação da gestão, com vistas a promover melhorias no desempenho de organizações, programas e atividades governamentais, quanto à eficiência, eficácia, economicidade e efetividade. Essa modalidade de avaliação está inserida, principalmente, nas atividades desenvolvidas pelos órgãos de controle interno e externo.

Além das avaliações governamentais apresentadas no QUADRO 26, de caráter geral, considerando que abrangem toda a administração pública federal, cabe acrescentar à tipologia apresentada, as avaliações instituídas no âmbito da política de avaliação e monitoramento do Ministério do Desenvolvimento Social e Combate à Fome - MDS, por intermédio da Portaria MDS n. 329, de 11 de outubro de 2006. Tal Portaria tem como diretriz contribuir para o aprimoramento da gestão pública, fornecendo elementos que contribuam para o aumento da responsabilização, eficiência, eficácia e efetividade das políticas sociais, e para o exercício do controle social sobre essas políticas. É baseada nos princípios da ética, da transparência, da imparcialidade, bem como da isenção e pautada pelos princípios de cooperação, articulação e integração entre os órgãos internos do MDS (BRASIL, 2006i).

De acordo com a mencionada Portaria, são realizados dois tipos de avaliação, monitoramento e avaliação, conforme apresentado no QUADRO 26.

Tipologia	Objeto (o que é)	Objetivo (para quê)
Monitoramento	Definir indicadores para programas e ações.	Obter informações para subsidiar o gerenciamento e a tomada de decisões cotidianas e identificar problemas.
	Definir dados a serem coletados para programas e ações.	
	Acompanhar a execução física e financeira dos programas e ações.	
	Construir e implantar sistemas de informações dos programas e ações.	
	Coletar e receber sistematicamente dados dos programas e ações.	
	Supervisionar <i>in loco</i> a execução dos programas e ações realizados pelo MDS e outros parceiros.	
	Produzir indicadores para os programas e ações.	
	Analisar os indicadores dos programas e ações.	
Avaliação	Analisar a implementação de programas e ações.	Contribuir para o aprimoramento da gestão pública, fornecendo elementos que contribuam para o aumento da responsabilização, eficiência, eficácia e efetividade das políticas sociais e para o exercício do controle social sobre as políticas públicas, bem como contribuir para o acompanhamento dos Programas do Plano Plurianual (PPA).
	Analisar os resultados dos programas e ações.	
	Analisar os impactos ou efeitos dos programas e ações.	
	Analisar a eficiência, equidade, eficácia ou efetividade de programas e ações.	
	Analisar o perfil dos beneficiários dos programas e ações.	
	Elaborar diagnósticos de perfil da demanda.	
	Avaliar a satisfação de beneficiários/usuários.	
	Avaliar a qualidade dos serviços prestados.	
Elaborar estudos de linha de base		

QUADRO 27 - TIPOLOGIA DAS AVALIAÇÕES REALIZADAS PELO MDS

4.1 AS AVALIAÇÕES DO PLANO PLURIANUAL 2004-2007 REALIZADAS PELA SPI/MPOG

Com a reforma orçamentária implementada a partir do Plano Plurianual 2000-2003, os orçamentos da União passaram a estruturar toda ação finalística do Governo Federal em programas orientados para a consecução de objetivos estratégicos estabelecidos no PPA (BRASIL, 1998).

Os programas foram definidos a partir da problematização de demandas e necessidades da sociedade. Cada programa articula um conjunto de ações, com metas físicas e financeiras definidas, que concorrem para solução do problema previamente identificado. Além disso, os programas são gerenciados diretamente por um responsável (gerente) e mensurados por indicadores de resultado.

Anualmente, os objetivos estratégicos do Governo e os resultados dos programas são avaliados para subsidiar a elaboração da Lei de Diretrizes Orçamentárias. Essas avaliações integram o Relatório de Avaliação do PPA, que é apresentado ao Congresso Nacional no exercício subsequente.

O Decreto n. 2.829/1998 estabeleceu que a avaliação física e financeira dos Programas, bem como dos projetos e atividades que os constituem, é inerente às responsabilidades da unidade responsável pelo programa e tem por finalidade:

- i) aferir o resultado, tendo como referência os objetivos e as metas fixadas;
- ii) subsidiar o processo de alocação de recursos públicos, a política de gastos públicos e a coordenação das ações de governo;
- iii) evitar a dispersão e o desperdício de recursos públicos. (BRASIL, 1998).

As leis que aprovam os planos plurianuais estabelecem os critérios para avaliação tanto do plano como dos programas.

Com a edição do Plano Plurianal 2004-2007, foi determinada a criação de um sistema de monitoramento e avaliação do PPA, cuja estrutura foi definida pelo Decreto n. 5.233 de 6 de outubro de 2004, que dispôs também sobre a gestão do Plano e dos Programas (BRASIL, 2004).

A gestão do PPA 2004-2007 foi orientada pelos critérios de eficiência, eficácia e efetividade, cujas definições apresentadas no Decreto 5.233/2004 são as seguintes:

- Eficiência é a medida da relação entre os recursos efetivamente utilizados para a realização de uma meta para um projeto, atividade ou programa em relação ao previsto.
- Eficácia é a medida do grau de atingimento das metas fixadas para um determinado projeto, atividade ou programa em relação ao previsto.

- Efetividade é a medida do grau de atingimento dos objetivos que orientaram a constituição de um determinado programa, expressa pela sua contribuição à variação alcançada dos indicadores estabelecidos no plano (BRASIL, 2004).

O sistema de monitoramento e avaliação tem por objetivo assegurar que a avaliação seja parte integrante da gestão dos programas e dê subsídios para a tomada de decisão, sendo formado pela Comissão de Monitoramento e Avaliação (CMA) e pelas Unidades de Monitoramento e Avaliação (UMA) de cada ministério setorial e secretaria especial (MPOG, 2004).

São objetivos específicos da avaliação, segundo o Manual de Avaliação do PPA 2004-2007: Ano Base 2004:

- i) proporcionar maior transparência às ações de governo;
- ii) auxiliar a tomada de decisão;
- iii) promover a aprendizagem e a disseminação do conhecimento nas organizações; e
- iv) aperfeiçoar a concepção e a gestão do plano e dos programas.(MPOG, 2004c)

Para dar cumprimento aos objetivos propostos, o processo de avaliação foi realizado por meio de questionário padronizado, autoaplicado em três etapas e instâncias: pelo gerente do programa, pelo ministério setorial e pelo Ministério do Planejamento, Orçamento e Gestão. Coube ao gerente a avaliação do programa quanto aos resultados, concepção e implementação; ao ministério setorial, a avaliação setorial, compreendendo os resultados, a concepção do conjunto de programas e a gestão. Quanto ao MPOG, a avaliação do Plano, no que diz respeito ao cenário macroeconômico, desafios e gestão do Plano.

A avaliação da concepção do programa teve por objetivo analisar os aspectos do desenho do programa, quanto à adequação do objetivo, caracterização e dimensionamento do público-alvo, regionalização, pertinência e suficiência das ações e das metas físicas, bem como adequação dos indicadores e da matriz de fontes de financiamento.

A avaliação da implementação teve por objetivo identificar os fatores que interferiram na obtenção dos resultados, no que diz respeito à suficiência e adequação dos meios para a implementação e o gerenciamento do programa; à disponibilidade de informações mínimas necessárias para permitir o controle da execução de todas as ações do programa, inclusive as ações executadas de forma descentralizada, as executadas em outras unidades administrativas ou mesmo outros ministérios; e, a análise das restrições que dificultaram o desempenho do programa (MPOG, 2004c).

A avaliação dos resultados diz respeito à evolução dos índices dos indicadores em relação aos objetivos dos programas.

Para o exercício de 2005, foi adotado novo manual avaliação o qual manteve a mesma estrutura de análise (concepção, implementação e resultados), assim como para os exercícios seguintes 2006 e 2007, com pequenas mudanças em relação às questões abordadas, como, por exemplo, a partir de 2006 foram incluídas perguntas sobre a implementação do programa relativas às questões de gênero e atendimento aos grupos vulneráveis, como deficientes, crianças e jovens. E, no exercício de 2007, foi incluída uma nova seção no questionário, que tratou da avaliação do processo de avaliação.

O PPA 2004-2007 inovou também quanto à flexibilização da sua rigidez, sendo criada a figura do PPA deslizante, permitindo a revisão anual do Plano, mediante a qual os programas podem ser modificados, alterados, descontinuados, criados ou até mesmo excluídos. De acordo com a Lei 10.933/2004, que aprovou o primeiro Plano para o período 2004-2007, são consideradas alterações no programa, os seguintes fatos:

- i) adequação de denominação ou do objetivo e modificação do público-alvo, dos indicadores ou dos índices;
- ii) inclusão ou exclusão de ações orçamentária;
- iii) alteração do tipo, do título, do produto, da unidade de medida e das metas das ações orçamentárias;
- iv) alteração dos valores estimados para cada ação, no período do Plano Plurianual, respeitada a respectiva regionalização (BRASIL, 2004).

Por outro lado, os programas excluídos a cada revisão do PPA constam de anexo específico. Desta forma, durante a vigência do Plano 2004-2007, o quantitativo de programas vigentes, que permaneceram com o mesmo código e denominação, assim como o quantitativo de programas avaliados, estão apresentados no TABELA 1.

TABELA 1 - COMPARATIVO DO QUANTITATIVO DE PROGRAMAS VIGENTES PPA 2004-2007 E AVALIADOS NO PERÍODO

Exercício	2004	%	2005	%	2006	%	2007	%
Número de Programas	394	100	393*	100	362*	100	358*	100
Número de Programas Avaliados	360	91	368	93	284	78	299	83

FONTE: Sigplan e Relatórios de Avaliação do PPA 2004-2007

NOTA: * foram excluídos desses quantitativos os programas que constam do Sigplan nesses exercícios, apenas com inscrição em Restos a Pagar (RAP), respectivamente, 15 programas em 2005, 31 em 2006 e 3 em 2007.

Conforme se pode observar, em que pese a obrigatoriedade de avaliação de todos os programas, no período examinado, não houve o pleno cumprimento desse dispositivo. Nos dois primeiros exercícios, a média de avaliação dos programas ficou acima dos 90%, sendo que, a partir de 2006, em decorrência da não avaliação dos programas de gestão, esse percentual ficou em torno dos 80%.

Durante a vigência do PPA 2004-2007, foram avaliados 419 programas, sendo que apenas 55% desse total foi avaliado durante os 4 anos de vigência do Plano.

Programas Avaliados	Número	%
4 anos	232	55
3 anos	56	13
2 anos	84	20
1 ano	47	11
Total de Programas Avaliados no Período 2004-2007	419	100

QUADRO 28 - COMPARATIVO DA FREQUÊNCIA DAS AVALIAÇÕES DOS PROGRAMAS NO PERÍODO VIGÊNCIA PPA 2004-2007

FONTE: RELATÓRIOS DE AVALIAÇÃO DO PPA 2004-2007

4.2 AS AVALIAÇÕES DA SECRETARIA FEDERAL DE CONTROLE INTERNO/CGU

Conforme dispõe a Instrução Normativa/SFC n. 1/2001,

“para o cumprimento de suas atribuições e alcance de suas finalidades, o Sistema de Controle Interno do Poder Executivo Federal deve realizar suas atividades com foco nos resultados da ação governamental, para o qual utilizará de metodologia de planejamento específica, com ênfase na visão dos programas de governo como fator básico de organização da função pública, e na gestão pública como mobilização organizacional para o alcance desses resultados (BRASIL, 2001b).

A execução dos programas constitui, portanto, objeto central de atuação do Sistema de Controle Interno, orientadora da estratégia de planejamento das ações de controle, que inicia com o mapeamento das políticas públicas de cada ministério, para identificação dos macro-objetivos, recursos, agentes responsáveis e interfaces, com o objetivo de conhecer a importância estratégica de cada uma dessas políticas e seus programas.

Após o mapeamento das políticas e identificação dos programas, estes são hierarquizados, segundo a percepção do Controle, em essenciais, relevantes e coadjuvantes, com base em critérios políticos e estratégicos, considerando os riscos relativos à materialidade, relevância e criticidade.

Uma vez selecionados os programas para atuação do Controle Interno, detalha-se as ações dos respectivos programas com vistas à elaboração de relatório de situação. Em seguida, é definido o plano estratégico da ação selecionada para investigação, incluindo o levantamento dos pontos críticos e frágeis com potencial para impactar a sua execução, e se define a abordagem a ser adotada. Com base no plano estratégico, são elaborados o plano operacional e os instrumentos de coleta de dados. Nessa ocasião também se estabelece a época mais oportuna para realização dos trabalhos.

A metodologia de trabalho adotada pela SFC, segundo a IN/SFC n. 01/2001, abrange

os aspectos pertinentes ao conhecimento das Ações de cada Ministério, à percepção sobre as principais áreas de atuação e pauta política específica, à compreensão sobre o funcionamento dos Programas e respectivas Ações ou equivalentes, bem como avaliação sobre as unidades responsáveis pelas diferentes etapas do processo gerencial de implementação da Política Pública e o planejamento das ações de controle a serem implementadas para avaliação da execução dos Programas de Governo e da Gestão Pública (BRASIL, 2001).

A estratégia adotada pela SFC classifica as ações de controle em sistemáticas, assistemáticas (sorteios) e operações especiais. O controle sistemático tem como premissa a existência do planejamento mencionado anteriormente e o assistemático diz respeito às excepcionalidades, que, pelas suas características e riscos, demandam a intervenção do Controle dispensando-se o planejamento completo.

As ações sistemáticas são selecionadas utilizando-se a Curva ABC, segundo os critérios de hierarquização já citados, onde são priorizados, em média, 60 programas por ano dentre aqueles constantes da LOA. A partir da identificação das principais ações e da elaboração do relatório de situação, há interação com o gestor responsável, para melhor compreender o programa e seus pontos críticos com vistas à elaboração do plano estratégico, com base no qual será elaborado o plano operacional, que detalha a forma de execução das ações de controle.

Utilizando-se amostragem estatística são selecionados os locais a serem visitados para aplicação dos instrumentos de coleta de dados (questionário estruturado e relatos de constatação).

Depois de concluída 35% da execução do plano operacional, elabora-se o primeiro balanço, por meio de nota técnica, detalhando a situação constatada por município e por programa. Essa nota técnica é discutida com os gestores responsáveis, onde são apresentadas recomendações para adoção de providências (caso geral). Nessa oportunidade, negocia-se com o gestor as medidas necessárias para implementar as recomendações sugeridas na nota técnica.

O acompanhamento da implementação das recomendações é realizado de acordo com a estratégia de cada Coordenação de Área da SFC, e, necessariamente, quando do levantamento de Contas Anuais, no caso das Unidades que apresentam Contas. No entanto, a SFC adota um plano permanente de providências para o acompanhamento contínuo e sistemático de tudo que foi produzido pela Secretaria.

As ações assistemáticas são realizadas por intermédio dos sorteios de municípios para fiscalização da aplicação dos recursos federais transferidos. Nessas ações, para cada município selecionado, de acordo com o número de habitantes, levantam-se os dados relativos às transferências de recursos federais realizadas. Para os municípios com até 20 mil habitantes, são identificados todos os recursos repassados; entre 20 e 100 mil, aqueles repassados para as áreas da saúde, desenvolvimento social, educação e mais uma área sorteada. Acima de 100 mil habitantes, procede-se ao levantamento dos dados relativos aos recursos transferidos para a área sorteada e mais uma área a ser sorteada entre saúde, desenvolvimento social e educação.

O trabalho de campo é realizado de acordo com a natureza do objeto a ser fiscalizado; quando cabível, utiliza-se um plano amostral para o levantamento dos dados.

As ações de controle assistemáticas (sorteios) são programadas com base em cronograma, que fixa os prazos para conclusão de cada etapa de trabalho. Ao final é produzido um balanço geral.

Antes do balanço geral, são gerados relatórios agregados por ministério, discriminando a situação dos programas executados em cada município. No entanto, após os trabalhos de campo, é elaborado relatório preliminar (informativo ao prefeito), que é encaminhado para o prefeito para que se manifeste sobre as constatações no prazo de 10 dias. Após análise dos comentários do prefeito, o relatório é concluído e enviado novamente para o prefeito, como também para a câmara de vereadores, o promotor da comarca, o Ministério Público Federal, TCU, para as Comissões de Fiscalização da Câmara e Senado, para o ministério responsável pelas áreas das ações fiscalizadas, para AGU e, sempre que necessário, para a Polícia Federal. A partir da expedição dos avisos pelo Ministro da CGU, os relatórios finais são disponibilizados no sítio da CGU, na internet.

Finalizado o relatório dos sorteios, são encaminhadas, para os ministérios responsáveis, as respectivas notas técnicas com recomendações para implementação das medidas corretivas ou aperfeiçoadoras. No entanto, nesses casos não há negociação das medidas com os gestores, diferentemente do que ocorre quando das ações sistemáticas.

No que diz respeito ao acompanhamento das medidas, este é realizado como nas ações sistemáticas. Além disso, de acordo com Acórdão proferido pelo TCU, após 120 dias da divulgação dos resultados dos sorteios, deve ser encaminhado àquela Corte, balanço das providências adotadas para as recomendações decorrentes do sorteio em questão, bem como sobre as pendências dos sorteios anteriores.

A fiscalização dos programas, no período de 2003 a 2006, foi focada nas ações de sorteio. No entanto, a partir de 2006, se começou a retomar as ações sistemáticas, tendo se consolidado em 2007. Em decorrência, o número de sorteios tem diminuído.

Quanto às operações especiais, trata-se de um tipo de ação que associa a atividade de controle, a forma de implementação da fraude. Este tipo de atividade tem origem tanto nas denúncias que são encaminhadas à CGU, como no conhecimento acumulado das demais ações do Controle. Também podem ser realizadas a pedido da Polícia Federal ou do Ministério Público Federal e, necessariamente, envolve a participação da Polícia Federal.

Para realização das avaliações, o Sistema de Controle Interno do Poder Executivo Federal, considera os dados e as informações disponíveis nos sistemas e registros governamentais, os quais são mensurados por meio de um conjunto de dados de indicadores e indicadores que evidenciam o desenvolvimento da gestão quanto à eficiência, eficácia, legalidade e economicidade. A IN/SFC n. 01/2001 adota as seguintes definições para esses termos:

- Legalidade: Consiste na aderência dos atos e fatos de gestão praticados, aos normativos legais e técnicos que regem os mesmos.
- Economicidade: Expressa variação positiva da relação custo/benefício, na qual busca-se a otimização dos resultados na escolha dos menores custos em relação aos maiores benefícios. Revela a atenção da gestão com o bom uso qualitativo dos recursos financeiros, por definição, escassos, desde a adequação da proposta orçamentária das metas a serem atingidas, passando pela coerência com respeito aos preços de mercado, o desenvolvimento de fontes alternativas de receita e a obtenção dos menores custos por produto gerado.
- Eficácia: É o grau de atingimento das metas fixadas para um determinado objeto de uma ação em relação ao previsto, em um determinado período.
- Eficiência: É a medida da relação entre os recursos efetivamente utilizados para a realização de uma meta, frente a padrões estabelecidos. Mede, então, a utilização dos recursos de que a unidade ou entidade dispõe para realizar um conjunto de ações e operações que visam atingir um propósito de trabalho previamente programado. A eficiência está associada ao uso dos recursos disponíveis em relação aos produtos e serviços finais elaborados (BRASIL, 2001b).

Os instrumentos de controle utilizados pela SFC para realização de suas atribuições são a auditoria e a fiscalização. A auditoria tem por objetivo avaliar a gestão pública por intermédio dos processos e resultados gerenciais, como também a aplicação de recursos públicos por entidades de direito público e privado, com vistas a comprovar a legalidade e legitimidade dos atos e fatos administrativos e avaliar os resultados alcançados quanto aos aspectos de eficiência, eficácia e economicidade da gestão orçamentária, financeira, patrimonial, operacional, contábil e finalística das unidades da administração pública.

As auditorias classificam-se em:

- i) Auditoria de Avaliação da Gestão: esse tipo de auditoria objetiva emitir opinião com vistas a certificar a regularidade das contas, verificar a execução de contratos, acordos, convênios ou ajustes, a probidade na aplicação dos dinheiros públicos e na guarda ou administração de valores e outros bens da União ou a ela confiados, compreendendo, entre outros, os seguintes aspectos: exame das peças que instruem os processos de tomada ou prestação de contas; exame da documentação comprobatória dos atos e fatos administrativos; verificação da eficiência dos sistemas de controles administrativo e contábil; verificação do cumprimento da legislação pertinente; e avaliação dos resultados operacionais e da execução dos programas de governo quanto à economicidade, eficiência e eficácia dos mesmos.
- ii) Auditoria de Acompanhamento da Gestão: realizada ao longo dos processos de gestão, com o objetivo de se atuar em tempo real sobre os atos efetivos e os efeitos potenciais positivos e negativos de uma unidade ou entidade federal, evidenciando melhorias e economias existentes no processo ou prevenindo gargalos ao desempenho da sua missão institucional.
- iii) Auditoria Contábil: compreende o exame dos registros e documentos e na coleta de informações e confirmações, mediante procedimentos específicos, pertinentes ao controle do patrimônio de uma unidade, entidade ou projeto. Objetivam obter elementos comprobatórios suficientes que permitam opinar se os registros contábeis foram efetuados de acordo com os princípios fundamentais de contabilidade e se as demonstrações deles originárias refletem, adequadamente, em seus aspectos mais relevantes, a situação econômico-financeira do patrimônio, os resultados do período administrativo examinado e as demais situações nelas demonstradas. Tem por objeto, também, verificar a efetividade e a aplicação de recursos externos, oriundos de agentes financeiros e organismos internacionais, por unidades ou entidades públicas executoras de projetos celebrados com aqueles organismos com vistas a emitir opinião sobre a adequação e fidedignidade das demonstrações financeiras.
- iv) Auditoria Operacional: consiste em avaliar as ações gerenciais e os procedimentos relacionados ao processo operacional, ou parte dele, das unidades ou entidades da administração pública federal, programas de governo, projetos, atividades, ou segmentos destes, com a finalidade de emitir uma opinião sobre a gestão quanto aos aspectos da eficiência, eficácia e economicidade, procurando auxiliar a administração na gerência e nos resultados, por meio de recomendações, que visem aprimorar os procedimentos, melhorar os controles e aumentar a responsabilidade gerencial. Este tipo de procedimento auditorial, consiste numa atividade de assessoramento ao gestor público, com vistas a aprimorar as práticas dos atos e fatos administrativos, sendo desenvolvida de forma tempestiva no contexto do setor público, atuando sobre a gestão, seus programas governamentais e sistemas informatizados.
- v) Auditoria Especial: objetiva o exame de fatos ou situações consideradas relevantes, de natureza incomum ou extraordinária, sendo realizadas para atender determinação expressa de autoridade competente. Classificam-se, nesse tipo, os demais trabalhos auditoriais não inseridos em outras classes de atividades (BRASIL, 2001b).

No que diz respeito às fiscalizações, trata-se de um instrumento de controle cuja finalidade é avaliar a execução dos programas, orçamentários ou não, quanto à execução das metas e objetivos estabelecidos e à qualidade do gerenciamento; bem como aferir a adequação dos mecanismos de controle social sobre os programas contemplados com recursos da União.

Pode-se observar que o resultado das ações de fiscalização e auditoria realizadas pela SFC é a avaliação. No entanto, a sistemática do controle e a abordagem metodológica são orientadas pelas normas e padrões de auditoria, conforme definidas na IN/SFC n. 01/2001:

- i) Normas Relativas aos servidores do Sistema de Controle Interno do Poder Executivo Federal.
- ii) Normas Relativas à execução do trabalho no Sistema de Controle Interno do Poder Executivo Federal.
- iii) Normas Relativas à opinião do Sistema de Controle Interno do Poder Executivo Federal.
- iv) Normas Relativas à forma de comunicação.
- v) Normas Relativas à audiência do auditado.
- vi) Normas Relativas às auditorias de recursos externos.
- vii) Normas Relativas ao controle interno administrativo (BRASIL, 2001b).

Desta forma, a SFC não utiliza o instrumento avaliação, segundo os padrões de avaliação, diferentemente do observado nos trabalhos de auditoria operacional realizados pelo TCU e apresentados a seguir, muito embora o produto das auditorias e fiscalizações por eles realizadas seja a avaliação dos programas de governo quanto aos seus aspectos legais e de desempenho, especialmente quanto à eficiência, a economicidade e a eficácia.

4.3 AS AVALIAÇÕES DE PROGRAMAS DO TRIBUNAL DE CONTAS DA UNIÃO REALIZADAS NO PERÍODO 2004-2007

A Constituição Federal de 1988, em seu Artigo 71, Inciso IV, dispõe que compete ao Tribunal de Contas da União, por iniciativa própria, da Câmara dos Deputados, do Senado Federal, de comissão técnica ou de inquérito, realizar inspeções e auditorias de natureza contábil, financeira, orçamentária, operacional e patrimonial, nas unidades administrativas dos Poderes Legislativo, Executivo e Judiciário, como também, nas unidades da administração direta e indireta, incluídas as fundações e sociedades instituídas e mantidas pelo poder público federal (BRASIL, 1988).

A Lei Orgânica do TCU, nº 8.443/92, além de reproduzir o mandamento constitucional no inciso II do Art.1º, dispõe, em seu Art. 3º, que assiste ao Tribunal o poder regulamentar no âmbito de sua competência e jurisdição. Assim, coube ao Regimento Interno (RITCU), aprovado pela Resolução nº 155, de 4 de dezembro de 2002, no capítulo III, seção II, ao tratar dos instrumentos de fiscalização, defini-los e classificá-los em: Levantamentos

(Art. 238); Auditorias (Art. 239); Inspeções (Art. 240); Acompanhamentos (Art. 241); e Monitoramentos (Art. 243) (BRASIL, 2002).

Auditoria, segundo o Regimento, é o instrumento de fiscalização utilizado pelo Tribunal para (Art. 239 RITCU):

- i) examinar a legalidade e a legitimidade dos atos de gestão dos responsáveis sujeitos a sua jurisdição, quanto ao aspecto contábil, financeiro, orçamentário e patrimonial;
- ii) avaliar o desempenho dos órgãos e entidades jurisdicionados, assim como dos sistemas, programas, projetos e atividades governamentais, quanto aos aspectos de economicidade, eficiência e eficácia dos atos praticados;
- iii) subsidiar a apreciação dos atos sujeitos a registro (BRASIL, 2002).

Para dar cumprimento aos normativos relacionados, especialmente quanto às auditorias destinadas a avaliar o desempenho, o Manual de Auditoria de Natureza Operacional – Anop - aprovado pela Portaria/TCU nº 144, de 10 de julho de 2000, definiu que esse tipo de fiscalização consiste na avaliação sistemática dos programas, projetos e atividades governamentais, assim como dos órgãos e entidades jurisdicionados ao Tribunal, no que se refere à sua gestão ou ao impacto de sua ação, quanto aos aspectos da economicidade, eficiência, eficácia e efetividade, que se referem a:

- Economicidade: minimização dos custos dos recursos utilizados na consecução de uma atividade, sem comprometimento dos padrões de qualidade.
- Efetividade: relação entre os resultados (impactos observados) e os objetivos (impactos esperados).
- Eficácia: grau de alcance das metas programadas em um determinado período de tempo, independentemente dos custos implicados.
- Eficiência: relação entre os produtos (bens e serviços) gerados por uma atividade e os custos dos insumos empregados em um determinado período de tempo. O resultado expressa o custo de uma unidade de produto final em um dado período de tempo. Analogamente ao conceito de eficácia, pode-se estabelecer o grau de eficiência relativa de um programa, projeto ou atividade comparando-o com as metas e custos programados ou com dados de outras iniciativas que perseguem os mesmos objetivos. Nesse caso, somente serão comparáveis iniciativas que, além dos objetivos, possuam as mesmas características de programação (BRASIL, 2000).

Para a realização das auditorias operacionais, o Tribunal observa as diretrizes estabelecidas pela *International Organization of Supreme Audit Institutions - Intosai* (2003), segundo a qual, a auditoria de desempenho, ou de gestão, refere-se ao exame da economia, eficiência e eficácia dos programas e organizações governamentais. Essas dimensões de desempenho são examinadas mediante questões que abordam, por exemplo, se a forma de utilização dos insumos representa o uso mais econômico dos recursos públicos; se estão sendo

disponibilizados os melhores serviços em relação aos recursos disponíveis; ou se os objetivos da política estão sendo atingidos e os impactos são resultantes dessa política.

Essa modalidade de auditoria tem foco no desempenho em detrimento dos valores realizados e contábeis, e são classificadas em auditoria de desempenho operacional e avaliação de programas (BRASIL, 2002).

Segundo a *Intosai*, em alguns países, as avaliações de programa são realizadas como uma modalidade de auditoria de desempenho, com o objetivo de examinar o desempenho dos programas e as alternativas para melhorar os seus resultados.

A avaliação é uma ferramenta utilizada para subsidiar a tomada de decisão tanto para o gestor, quanto para a esfera política e serve ao princípio democrático da prestação de contas. O papel do controle externo da Administração é o de garantir a fidedignidade das informações produzidas pelas avaliações dos gestores ou mesmo, quando necessário, na produção de dados sobre desempenho. E assim, contribuir para a responsabilização dos gestores públicos pelo desempenho alcançado (SERPA, 2004).

A auditoria de natureza operacional, portanto, se constitui em um processo de coleta e análise sistemáticas de informações sobre características, processos e resultados de um programa, atividade ou organização, com base em critérios fundamentados, com o objetivo de subsidiar os mecanismos de responsabilização por desempenho e contribuir para aperfeiçoar a gestão pública (BRASIL, 2004b).

Responsabilização por desempenho é uma das dimensões do paradigma de *accountability*, conforme se depreende do conceito apresentado pelo Grupo de Pesquisa Tendências do Controle (1999), pois se relaciona à noção de governança, à capacidade do governo de responder às demandas da sociedade, à transparência das ações do poder público e à responsabilidade dos agentes políticos e administradores públicos pelos seus atos, transcendendo-se, assim, ao conceito de prestação de contas tradicional (SERPA, 2004).

O papel do Controle Externo é assegurar a *accountability* tanto nas dimensões de legalidade/conformidade, como na de resultado/desempenho. Desta forma, a auditoria de natureza operacional é um dos instrumentos a serem utilizados na construção dos mecanismos que possibilitem a responsabilização dos gestores públicos por desempenho (SERPA, 2004).

Outro objetivo almejado é contribuir para melhoria da gestão, que se materializa mediante a formulação de recomendações aperfeiçoadoras.

As auditorias de natureza operacional são realizadas segundo um ciclo de atividades, que tem início com a seleção do objeto a ser investigado (programa, ação, projeto, atividade, órgãos/entidades, sistemas, dentre outros), mediante a aplicação de matriz de risco e a

realização do estudo de viabilidade. Uma vez identificado o objeto da auditoria, passa-se à fase do planejamento, seguida pela execução, elaboração do relatório e encaminhamento deste para os comentários do gestor. Após analisados os comentários do gestor, o relatório é submetido à apreciação do Tribunal. Procede-se a divulgação dos resultados e, posteriormente, ao monitoramento das recomendações formuladas pelo TCU (SERPA e MARTINS, 2005).

A fase de seleção, primeira etapa do ciclo de trabalhos de Anop, é processada por meio de Levantamentos, conforme definido no Regimento Interno do TCU (Brasil, 2002), que têm como objetivos conhecer a forma de funcionamento e coletar informações sobre a execução do programa, com vistas a avaliar a oportunidade e viabilidade de se realizar o trabalho.

Cabe observar que, segundo as diretrizes para auditorias de desempenho da *Intosai*, a estratégia metodológica a ser utilizada para seleção de temas (ou áreas) de auditoria, deve considerar as seguintes etapas:

- determinar as possíveis áreas de fiscalização a partir das quais se selecionarão as áreas estratégicas para atuação da instituição - implica escolhas estratégicas considerando o número de áreas, a possibilidade de ampliação e a capacidade limitada da EFS; e
- estabelecer os critérios que serão empregados para a seleção das áreas - provavelmente o principal critério para seleção seja a potencial contribuição da auditoria para a avaliação e a melhoria do funcionamento da administração pública e dos organismos a ela vinculados (INTOSAI, 2003, tradução nossa).

Os critérios gerais sugeridos pela *Intosai* para a seleção de áreas (temas) para auditoria de desempenho são:

- valor agregado: quanto maiores as perspectivas de se realizar uma fiscalização útil, de qualidade elevada e quanto menor o número de trabalhos de avaliação realizados no tema em questão, maior será o valor agregado correspondente.
- problemas ou áreas com problemas importantes: quanto maior for o risco de comprometimento da economia, eficiência e eficácia ou da credibilidade do setor público, mais importantes tendem a ser os problemas. Talvez este seja o critério mais relevante para o enfoque orientado para problemas;
- riscos ou incertezas: quanto maior for o interesse público envolvido quando existam suspeitas de ineficiências, maiores serão os riscos, e quanto menor o conhecimento a respeito, maior a incerteza (INTOSAI, 2003, tradução nossa).

Nessa fase, portanto, examinam-se as prioridades de governo, a relevância das ações para reduzir os problemas da sociedade, a materialidade dos recursos envolvidos e o risco de não se alcançarem os resultados esperados. Essas informações são pontuadas em uma Matriz de Riscos, segundo atributos definidos para os critérios: relevância, materialidade e risco, a

fim de se obter uma classificação em ordem decrescente. Os programas mais bem pontuados são indicados para elaboração de estudos de viabilidade.

Os estudos de viabilidade indicam a oportunidade e a modalidade de fiscalização mais adequada à auditoria, considerando os potenciais problemas que comprometem o desempenho.

Cabe observar que, a partir do exercício de 2007, o TCU passou a selecionar os trabalhos de auditoria operacional com base em estudos dos Temas de Maior Significância (TMS), não mais aplicando a matriz de risco.

A segunda fase do ciclo de trabalhos de ANOp - Planejamento, destina-se à definição do escopo da auditoria, bem como da elaboração dos procedimentos a serem utilizados para coleta de dados. Nesta fase, define-se o projeto de auditoria que é apresentado no instrumento denominado Matriz de Planejamento, que acompanha o relatório de planejamento.

A metodologia utilizada pelo Tribunal para os trabalhos de ANOp está estruturada a partir da identificação de problemas que comprometem o desempenho dos programas e/ou órgãos e entidades. Os problemas são diagnosticados por intermédio da aplicação de algumas técnicas, como por exemplo: mapa de processo, mapa de produto, análise stakeholder, análise RECI, análise de controles internos, análise SWOT e diagrama de verificação de risco.

O problema se constitui no escopo da auditoria e é examinado a partir de questões que objetivam averiguar o resultado das ações implementadas. As questões relacionam o problema às dimensões examinadas nos trabalhos dessa natureza, ou seja: economicidade, eficiência, eficácia, efetividade e equidade.

A Matriz de Planejamento apresenta, para cada questão, as informações que serão necessárias para análise, bem como a sua fonte, a estratégica metodológica para coleta e análise dessas informações, as limitações dela decorrente e os potenciais resultados esperados, ou seja, o que a análise vai permitir dizer.

Na fase de execução, a equipe de auditoria realiza os trabalhos de campo e as pesquisas necessárias à coleta de dados - por meio de entrevistas, aplicação de questionários, observação direta, grupos focais, consultas a documentos e bases de dados.

Após os trabalhos de campo, a equipe elabora a Matriz de Achados, síntese dos resultados obtidos nessa fase.

Cabe observar que tanto a Matriz de Planejamento como a Matriz de Achados são apresentadas e discutidas nos Painéis de Referência realizados pelo Tribunal com a participação de representados de grupos de interesse, como Controle Interno, Sociedade Civil, Academia e Consultorias Legislativas do Congresso. Nessas ocasiões, busca-se verificar a

consistência da proposta de auditoria, bem como a relevância e adequação das conclusões (achados de auditoria) e propostas de recomendações. As sugestões dos participantes são incorporadas à versão final das respectivas matrizes e apresentadas à consideração dos gestores dos programas.

Concluída a análise dos dados, elabora-se o relatório preliminar, a ser submetido ao gestor público para comentários. O relatório final representa o resultado do trabalho após a análise dos comentários, o qual é submetido às considerações do Ministro Relator para apreciação pelo Tribunal. Uma vez proferida a decisão do TCU, dá-se ampla publicidade.

A etapa de divulgação do relatório tem a finalidade de ampliar o conhecimento da sociedade sobre os resultados das ações estatais avaliadas, contribuindo para aumentar a efetividade do controle, por meio da mobilização da comunidade no acompanhamento e na apreciação dos objetivos, da implementação e dos resultados das políticas públicas.

Para assegurar que as recomendações formuladas pelo Tribunal sejam implementadas, procede-se ao monitoramento das recomendações durante, em média, um período de dois anos.

No período 2004-2007, foram realizados 36 trabalhos de auditoria operacional, na modalidade avaliação, coordenadas ou com a participação da Secretaria de Fiscalização e Avaliação de Programas do TCU. Cabe observar, no entanto, que no exercício de 2007, conforme mencionado, o processo de seleção dos trabalhos foi modificado, passando a serem utilizados os estudos dos Temas de Maior Significância (TMS). No QUADRO 29, apresenta-se, por exercício, a relação dos trabalhos de avaliação realizados pela ou com a participação da Seprog/TCU.

ANO	AVALIAÇÕES	TOTAL
2004	Programa Fome Zero	9
	Ação Transferência de Renda com Condiionalidades - Bolsa Família	
	Programa Combate à Exploração Sexual de Crianças e Adolescentes	
	Ação Assistência Financeira para Aquisição e Distribuição de Medicamentos Excepcionais	
	Programa Banco de Alimentos	
	Ações de Atenção à Saúde Mental	
	Programa Assistência a Vítimas e Testemunhas Ameaçadas	
	Programa Sistema Único de Segurança Pública	
	Programa Assistência Jurídica Integral e Gratuita	

(continua)

(conclusão)

ANO	AVALIAÇÕES	TOTAL
2005	Programa Segurança Fitozoossanitária no Trânsito de Produtos Agropecuários	11
	Programa de Desenvolvimento da Bovideocultura	
	Ação Construção de Cisternas para o Armazenamento de Água	
	Programa Segundo Tempo	
	Ações Pesquisa sobre Populações Indígenas e Fiscalização de Terras Indígenas	
	Programa Nacional de Eliminação da Hanseníase	
	Programa Doação Captação e Transplante de Órgãos e Tecidos	
	Ação Orientação Profissional e Intermediação de Mão-de-Obra	
	Programa Manutenção de Hidrovias	
	Ação Apoio à Formação Continuada de Professores do Ensino Fundamental	
	Programa Governo Eletrônico	
2006	Programa Desenvolvimento Sustentável da Aqüicultura	8
	Ação Levantamentos Hidrogeológicos do Programa Geologia do Brasil	
	Ação Atenção aos Pacientes Portadores de Coagulopatias	
	Programa Resíduos Sólidos	
	Resolução de Controvérsias no âmbito do MERCOSUL	
	Vigilância Sanitária de Produtos, Serviços e Ambientes	
	Ação de Transferência de Renda com Condicionaisidades - Bolsa Família (Cadastro)	
	Aquífero Guarani	
2007	Fundação Nacional de Saúde	8
	Sistema e Promoção Comercial do Ministério das Relações Exteriores	
	Estrutura e Ações Voltadas ao Controle de Tráfego Aéreo	
	Programa Nacional de Atividades Nucleares	
	Financiamento da Segurança Pública	
	Processo de Licenciamento Ambiental	
	Integração de Políticas e Ações Voltadas para a Segurança Pública	
	Avaliação do Impacto das Perdas Técnicas e Não-Técnicas em todo o Sistema Elétrico	

QUADRO 29 - AVALIAÇÕES REALIZADAS PELO TCU NO PERÍODO 2004-2007

FONTE: Seprog/TCU

4.4 AS AVALIAÇÕES REALIZADAS PELA SECRETARIA DE AVALIAÇÃO E GESTÃO DA INFORMAÇÃO - SAGE/MDS

Por intermédio da Portaria 329/2006, foi instituído no Ministério de Desenvolvimento e Combate a Fome a Política de Monitoramento e Avaliação, que tem como diretriz contribuir para o aprimoramento da gestão pública, fornecendo elementos que contribuam para o aumento da responsabilização, eficiência, eficácia e efetividade das políticas sociais, e para o exercício do controle social sobre essas políticas. É baseada nos princípios da ética, da transparência, da imparcialidade, bem como da isenção e pautada pelos princípios de cooperação, articulação e integração entre os órgãos internos do MDS.

A execução da Política é assegurada, dentre outros meios, pelo conjunto dos seguintes instrumentos (BRASIL, 2006i):

- i) plano anual de monitoramento e avaliação;
- ii) sistema de monitoramento;
- iii) relatório de acompanhamento da execução física e financeira;
- iv) relatórios das avaliações específicas;
- v) relatórios semestrais dos indicadores de monitoramento;
- vi) relatório anual de financiamento da assistência social no Brasil; e
- vii) relatório anual das ações de monitoramento e avaliação.

A Política é implementada por meio das ações de monitoramento e avaliação. São consideradas ações de monitoramento aquelas que se destinam ao acompanhamento da implementação e execução dos programas e ações, visando à obtenção de informações para subsidiar o gerenciamento e a tomada de decisões cotidianas, bem como a identificação precoce de eventuais problemas, por meio, dentre outras, das seguintes atividades (BRASIL, 2006i):

- i) definição do conjunto mínimo de indicadores que devem ser monitorados por cada programa ou ação;
- ii) definição do conjunto mínimo de dados que devem ser coletados por cada programa ou ação;
- iii) acompanhamento sistemático da execução física e financeira dos programas e ações;
- iv) construção e implantação de sistemas de informações dos programas e ações;
- v) coleta ou recebimento sistemático dos dados referentes aos programas e ações;
- vi) supervisão *in loco* dos programas e ações executados pelo MDS ou por seus parceiros governamentais e não-governamentais;
- vii) produção de indicadores referentes aos programas e ações; e
- viii) análise dos indicadores referentes aos programas e ações.

No que diz respeito às ações de avaliação, é considerado qualquer estudo ou pesquisa referente aos programas e às ações que tenham um dos seguintes objetivos (BRASIL, 2006i):

- i) análise da implementação de programas e ações;
- ii) análise de resultados imediatos dos programas e ações;
- iii) análise de impactos ou efeitos dos programas e ações;
- iv) análise da eficiência, da equidade, da eficácia ou da efetividade de programas e ações;
- v) análise do perfil dos beneficiários dos programas e ações;

- vi) elaboração de diagnósticos de perfil da demanda;
- vii) avaliação da satisfação de beneficiários/usuários;
- viii) avaliação da qualidade dos serviços prestados; e
- ix) elaboração de estudos de “linha de base”.

As ações de monitoramento e avaliação devem contribuir para o acompanhamento dos Programas do Plano Plurianual (PPA).

A Secretaria de Avaliação e Gestão da Informação – SAGI - é o órgão responsável pela coordenação da Política de Monitoramento e Avaliação e dispõe de trabalhos de monitoramento e avaliação em diversos programas sob a responsabilidade do MDS, conforme apresentado no QUADRO 30.

Objeto da Avaliação	Objetivo
Programa Bolsa Família	Avaliação do impacto do Programa Bolsa Família
	Pesquisa domiciliar com os beneficiários do Programa Bolsa Família
	O Programa Bolsa Família e o enfrentamento das desigualdades de gênero
	Pesquisa de percepção dos beneficiários do Programa Bolsa Família sobre condições de segurança alimentar e nutricional
	Estudo sobre a importância das transferências do Programa Bolsa Família na renda municipal
Benefício de Prestação Continuada – BPC	Estudo sobre a importância das transferências do Benefício de Prestação Continuada na renda municipal
	Estudo sobre o Benefício de Prestação Continuada: aspectos de demanda, cobertura e relação com o sistema previdenciário brasileiro
	Avaliação da implementação do Benefício de Prestação Continuada e dos resultados sobre os beneficiários
Serviços de Proteção Social às Crianças e aos Adolescentes Vítimas de Violência, Abuso e Exploração Sexual - Programa Sentinela	Estudo para elaboração do Índice de Elegibilidade dos Municípios (IEMS) ao Programa Sentinela
	Estudo qualitativo sobre o Enfrentamento ao Abuso e à Exploração Sexual de Crianças e Adolescentes - Programa Sentinela
Programa de Erradicação do Trabalho Infantil - PETI	Levantamento de beneficiários do Programa de Erradicação do Trabalho Infantil
População em Situação de Rua	Pesquisa realizada no I Encontro Nacional de População em Situação de Rua
Programa de Aquisição de Alimentos - PAA	Estudo comparativo das diferentes modalidades do PAA-Região Nordeste
	Estudo do impacto do PAA sobre os arranjos econômicos locais nas regiões nordeste e sul do Brasil
Programa de Aquisição de Alimentos - PAA Leite	Estudo sobre os beneficiários do PAA Leite: produtores e consumidores
Programa Cisternas do MDS	Avaliação ambiental da performance do Programa Cisternas do MDS em parceria com a ASA: Índice de Sustentabilidade Ambiental (ISA)
	Avaliação de impacto social e dos processos de seleção e capacitação do Programa Cisternas do MDS/PIMC-ASA
Carteira Indígena	Avaliação do Projeto Carteira Indígena: Segurança Alimentar e Desenvolvimento Sustentável em Comunidades Indígenas
Programa de Apoio a Comunidades Quilombolas	Pesquisa de avaliação do Programa de Ações Estruturantes nas Comunidades Remanescentes de Quilombos

(continua)

(conclusão)

Objeto da Avaliação	Objetivo
Restaurantes Populares	Levantamento do perfil dos usuários de restaurantes populares Mapeamento e caracterização de restaurantes populares
Cozinhas Comunitárias	Avaliação do Programa Cozinhas Comunitárias
Agricultura Urbana	Avaliação do Projeto Hortas Comunitárias
Banco de Alimentos	Avaliação do Programa Banco de Alimentos
Chamada Nutricional em crianças menores de cinco anos, em populações específicas, em situação de vulnerabilidade social no semi-árido e assentamentos rurais no nordeste	Suplemento Especial sobre Segurança Alimentar da Pesquisa Nacional por Amostra de Domicílios - PNAD 2004
Pesquisas de Opinião	Pesquisa de opinião sobre os programas do MDS Pesquisa de opinião pública para avaliar as políticas e a imagem do MDS

QUADRO 30 - SÍNTESE DAS PESQUISAS DE AVALIAÇÃO DE PROGRAMAS SOCIAIS DO MDS

FONTE: MDS

Os trabalhos avaliativos realizados pela SAGI/MDS são executados por instituições contratadas por intermédio de Acordos de Empréstimos Internacionais, que preveem, dentre seus componentes, a avaliação.

A equipe de profissionais da SAGI destaca-se pela sua alta qualificação, sendo composta por 42 profissionais, todos com graduação, sendo 3 especialistas, 12 mestres e 7 doutores. Ao longo de sua existência, a Secretaria vem desenvolvendo uma *expertise* na contratação de avaliações externas (VAITSMAN *et al*, 2006).

É importante observar que não obstante a SAGI/MDS venha desenvolvendo e implementando os instrumentos de monitoramento e avaliação para o Ministério, a sustentabilidade desses processos está diretamente relacionada à construção da capacidade de avaliação do corpo funcional. Nesse ponto, observou-se que a equipe da Secretária é formada por profissionais contratados para funções de confiança, terceirizados e outros com contratos decorrentes do programa de seleção simplificado, cujo prazo de vigência é determinado, estando próximo o seu término. Segundo informações obtidas em entrevistas realizadas com servidores da Secretaria, apenas dois servidores (gestores) são funcionários efetivos do governo federal. Essa constatação é relevante, porque pode comprometer a continuidade das atividades da própria Secretaria.

A partir do levantamento de informações sobre as sistemáticas e abordagens metodológicas aplicadas às avaliações governamentais, apresentadas neste capítulo, passaremos, nos capítulos 5 e 6 a seguir, ao exame da qualidade e credibilidade das avaliações do PPA 2004-2007 e das avaliações realizadas pelo TCU, com vistas a verificar em que medida essas avaliações foram concebidas e implementadas de forma a favorecer a utilização

dos resultados por elas produzidos. Serão utilizados, para tanto, os padrões de avaliação definidos pelo Comitê Conjunto de Padrões para Avaliação de Programas Educacionais, cujos dados foram coletados com base no questionário constante do Apêndice.

5 ANÁLISE DAS AVALIAÇÕES DE PROGRAMAS DO PLANO PLURIANUAL 2004-2007 REALIZADAS PELA SECRETARIA DE PLANEJAMENTO E INVESTIMENTOS ESTRATÉGICOS DO MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO - SPI/MPOG

As avaliações dos programas do plano plurianual passaram a ser, sistematicamente, realizadas pela SPI/MPOG a partir do PPA 2000-2003, quando da reforma dos orçamentos federais, com a adoção do modelo gerencial de planejamento e orçamento. Naquela oportunidade, Garcia (2001, p. 4) ressaltara que “Implantar um sistema de avaliação para o planejamento e a gestão governamentais, no entanto, não é algo trivial. Inexiste, na administração pública brasileira, uma prática consagrada ou uma cultura institucional de avaliação.”

Em que pese tal deficiência, o Decreto n. 2.829/1998, ao estabelecer a nova estrutura para organização dos gastos públicos, introduziu a obrigatoriedade da avaliação como instrumento para subsidiar a melhoria na alocação dos recursos públicos, bem como a política e a coordenação dos gastos governamentais.

Desta forma, criou-se na SPI/MPOG um sistema de monitoramento e avaliação responsável por definir, regulamentar, coordenar e supervisionar a atividade avaliativa no âmbito da administração pública federal.

Calmon e Gusso (2003), ao analisarem a experiência brasileira com a avaliação do PPA 2000-2003, observaram avanços no movimento pela implantação da avaliação e melhoria da gestão. Dentre os avanços, destacaram:

- i) o incentivo à introdução de uma cultura de avaliação, a partir do desenvolvimento e disseminação de práticas de monitoramento da gestão pública;
- ii) a melhor visibilidade das ações do governo por meio da obrigatoriedade do registro de dados e informações sobre o resultado dos programas;
- iii) como instrumento de *feedback* do governo no processo de revisão, negociação e composição anual do orçamento, bem como na formulação do PPA 2004-2007;
- iv) a possibilidade de identificação dos principais obstáculos ao bom desempenho, a partir da sistematização, gerenciamento e acompanhamento dos programas; e, ainda,
- v) o desenvolvimento do Sigplan possibilitou medir a interação entre os atores envolvidos no processo de avaliação.

Por outro lado, na percepção desses mesmos autores, também foram observados alguns entraves decorrentes tanto de fatores internos, relacionados à sistemática de avaliação,

à metodologia e ao modelo de gestão do PPA; como também de fatores externos, referentes às restrições orçamentárias e fiscais (Calmon e Gusso, 2003).

As constatações de Calmon e Gusso (2003), especialmente quanto aos entraves relativos à sistemática de avaliação, dizem respeito ao processo e procedimentos (metodologia) utilizados na avaliação dos programas do PPA, que dificultam a compreensão dos seus resultados e o alcance dos seus propósitos.

A literatura tem demonstrado que os propósitos das avaliações condicionam a utilização dela esperada. Para tanto, a forma como são concebidas e conduzidas determinam a sua utilidade. Nesse sentido, a observância dos padrões de avaliação serve como direcionador para assegurar a qualidade e credibilidade dos trabalhos avaliativos, favorecendo a sua utilização.

Para Calmon e Gasso (2003, p. 31), “a sistemática de avaliação do PPA constitui muito mais uma prática de monitoramento do que a consolidação de diversas avaliações.”

De fato, ao se examinar a sistemática de avaliação do PPA depreende-se que suas características são similares ao monitoramento, conforme definido por Poister (2004):

Sistemas de monitoramento de desempenho são concebidos para o acompanhamento, em intervalos regulares, de indicadores de programas, agências ou sistemas de desempenho, cujas informações são reportadas para os gerentes e outros públicos especificados. Seu propósito é fornecer informações objetivas tanto para os gestores, como para os políticos, com vistas a contribuir para o aperfeiçoamento das decisões e, assim, fortalecer o desempenho; como também, para promover a *accountability* para um conjunto de stakeholders, como os dos níveis mais altos de gerenciamento, agências centrais, conselhos de governo, agências de financiamento, organizações credenciadas, clientes e consumidores, grupos de interesse e o público em geral (Epstein, 1984; Wholey e Hatry, 1992; Broom, Caudle, Jennings e Newcomer, 2002; Hendricks, 2002) apud (Poister, 2004, p. 99, tradução nossa).

As avaliações do PPA objetivam auxiliar a tomada de decisão, promover o aprendizado e a disseminação do conhecimento nas organizações, aperfeiçoar a concepção e gestão do plano, bem como promover a *accountability* das ações do governo.

A SPI/MPOG, para realização dessa atividade, adota um modelo único para avaliação dos programas do PPA, operacionalizada por meio de questionário padronizado, autoaplicável, cabendo ao gerente responder às questões sobre a concepção, implementação e resultados dos programas sob sua responsabilidade.

Cabe destacar, no entanto, que a avaliação do Plano Plurianual é composta, além das avaliações dos programas especificamente, da avaliação setorial, onde o ministério competente avalia os resultados e a gestão sob a perspectiva do conjunto de programas de sua

competência; e da avaliação do cenário macroeconômico, dos desafios e da gestão do Plano como um todo, realizada pela SPI/MPOG.

Para o desenvolvimento deste trabalho, selecionaram-se apenas as avaliações dos programas, considerando que esses instrumentos são a base para o alcance dos objetivos propostos quando da institucionalização da avaliação do PPA.

Para realização dessas avaliações, tem-se mantido a mesma estrutura conceitual e metodológica desde a sua origem (2000), sendo que, para o PPA 2004-2007, foi introduzido um modelo de gestão, por meio do Decreto 5.233/2004, que definiu critérios orientadores para gestão do Plano: eficiência, eficácia e efetividade, conforme já apresentado no capítulo 4, tendo-se criado, ainda, um sistema de monitoramento e avaliação, com vistas a assegurar o cumprimento dos objetivos propostos.

A eficiência da gestão diz respeito ao grau de execução financeira, pois é a relação entre os recursos efetivamente gastos, ou seja, pagos, e aqueles previstos para realização de uma meta. Desta forma, o programa será mais eficiente na medida em tiver a capacidade de executar financeiramente o orçamento aprovado.

Quanto à eficácia, esse critério permite mensurar o percentual de alcance das metas aprovadas.

A efetividade trata do cumprimento dos objetivos, que pode ser expressa pela variação alcançada nos indicadores de resultado estabelecidos no Plano.

As definições dos critérios, adotados pela SPI/MPOG, evidenciam a importância da mensuração das metas físicas e financeiras para a gestão, com vistas a assegurar o cumprimento dos objetivos dos programas.

Essa constatação pode ser corroborada com a leitura da tipologia de avaliação apresentada no QUADRO 26, onde a avaliação das metas físicas e financeiras é um dos principais objetivos das avaliações governamentais. Portanto, a verificação da adequação e suficiência dessas variáveis é um ponto central e crítico nas avaliações dos programas do PPA.

Para tanto, na avaliação da concepção do programa, são requeridas informações sobre o desenho do programa, que dizem respeito à adequação do objetivo, caracterização e dimensionamento do público-alvo, regionalização, pertinência e suficiência das ações e das metas físicas, bem como adequação dos indicadores e da matriz de fontes de financiamento.

Já na avaliação da implementação, são identificados os fatores que interferem na obtenção dos resultados, especialmente quanto à suficiência e adequação dos meios para a implementação e o gerenciamento do programa; a disponibilidade de informações mínimas necessárias para permitir o controle da execução de todas as ações do programa, inclusive as ações executadas de forma descentralizada, as executadas em outras unidades administrativas

ou mesmo outros ministérios; e, a análise das restrições que dificultam o desempenho do programa.

Na avaliação dos resultados, as questões são direcionadas para levantar informações sobre a evolução dos índices dos indicadores dos programas e a perspectiva de se alcançar os resultados esperados e cumprir as metas aprovadas.

5.1 A ADEQUAÇÃO DAS AVALIAÇÕES DE PROGRAMAS DO PPA 2004-2007 AOS PADRÕES DE AVALIAÇÃO DE PROGRAMAS DO COMITÊ CONJUNTO DE PADRÕES PARA AVALIAÇÃO EDUCACIONAL DOS ESTADOS UNIDOS

Cabe observar, inicialmente, que não existem no Brasil padrões orientadores para a atividade avaliativa, conforme já mencionado no item 2.2, no entanto, considerando que os padrões estabelecidos pelo Comitê Conjunto foram instituídos a partir do exercício dessa atividade, seus fundamentos são inerentes à prática da avaliação e amplamente validados na teoria avaliativa. Desta forma, mesmo não institucionalizados formalmente, os princípios a serem observados quando da realização desse tipo de trabalho são tacitamente aceitos.

Espera-se, com a aplicação dos padrões, examinar em que medida as avaliações governamentais são consistentes com os fundamentos da prática avaliativa, de forma a garantir a sua credibilidade e qualidade, fatores que favorecem a utilização dos seus resultados.

As avaliações, como instrumentos práticos, segundo Weiss (1998), seguem uma racionalidade, sendo a utilidade o atributo estruturante desses trabalhos. Desta forma, a utilidade como principal atributo da avaliação, pode ser examinada a partir dos padrões que tratam do objeto avaliado (o quê), do objetivo da avaliação (para quê), do avaliador e dos destinatários da avaliação (para quem).

Calmon e Gusso (2003), ao examinarem a experiência de avaliação do Plano Plurianual 2000-2003 do Governo Federal, constaram que,

Em relação aos entraves internos inerentes à sistemática de avaliação, pode-se mencionar, primordialmente, o elevado grau de subjetividade observado nas avaliações. A avaliação do PPA é interna e os programas são avaliados pelos próprios gerentes responsáveis pela sua implementação, muitas vezes com base em procedimentos subjetivos e informações imprecisas. Ademais, os critérios adotados para interpretar as escalas utilizadas nos questionários de avaliação não são evidentes, e não se sabe, por exemplo, se as respostas representam a percepção pessoal do gerente, se correspondem aos resultados de discussões entre dirigentes e técnicos envolvidos na implementação do programa, ou, ainda, se incorporam dados ou resultados de análises ou avaliações realizadas externamente.

Esse problema associa-se ao gerado pela padronização da metodologia de avaliação. As perguntas de avaliação são as mesmas para todos os programas, e essa aplicação

indiscriminada faz que os questionários se tornem demasiado genéricos e as técnicas de avaliação adotadas desconsiderem as particularidades intrínsecas e inerentes às realidades de cada um dos programas. Nesse sentido, a sistemática de avaliação do PPA, na verdade, constitui muito mais uma prática de monitoramento do que a consolidação de diversas avaliações, como vem ocorrendo em alguns países (CALMON e GUSSO, 2003, p. 31).

Essas constatações, muito embora retratem as inconsistências das avaliações relativas ao PPA anterior, ainda são bastante apropriadas para as avaliações de programas do Plano 2004-2007.

A lógica adotada na sistemática de avaliação dos programas do PPA é paradoxal, pois parte das necessidades de informações da SPI/MPOG, que desempenha o papel de avaliador e também de quem define as questões a serem respondidas, sem considerar as necessidades dos atores envolvidos no processo, seus papéis, tampouco os problemas existentes que comprometem o desempenho e os resultados dos programas avaliados. Por outro lado, as questões propostas pela Secretaria objetivam identificar problemas relativos à concepção, implementação e riscos para o alcance das metas e dos índices estabelecidos para os indicadores dos Programas, com base nas informações declaradas pelos gerentes nas respostas ao questionário.

O relatório de avaliação dos programas apresentado anualmente ao Congresso é uma síntese das respostas oferecidas pelos gerentes ao questionário. Os dados e informações apresentados pelos gestores não são analisados e tratados, sequer são examinados à luz dos critérios definidos para gestão dos programas: eficiência, eficácia e efetividade. Além disso, em que pese os relatórios destinarem-se a *stakeholders* com diferentes interesses, não se consideram as peculiaridades dos diversos públicos, sendo o relatório apresentado em um único formato.

Da análise das respostas ao questionário, observou-se que não há um retorno sistemático da SPI/MPOG para os gerentes sobre as respostas encaminhadas, sobre os resultados e as deficiências apontadas. Além disso, se por um lado essas avaliações permitem identificar, sistematizar e dimensionar os problemas relativos às deficiências que dificultam a implementação dos programas e o alcance dos resultados, não há nenhuma evidência de quais medidas foram adotadas para investigar as causas desses problemas e propor soluções.

Como exemplo, pode-se depreender, da análise das avaliações examinadas dos programas gerenciados pelas unidades da administração direta (17 entidades), a reiterada identificação da inadequação e insuficiência dos recursos materiais e de infraestrutura, bem como de pessoal. Anualmente, foram oferecidas respostas afirmativas para as questões que

tratavam desses pontos, – relatadas nos relatórios de avaliação – mas para as quais não foram propostas medidas corretivas ou qualquer outra alternativa para sanar as deficiências.

Outra evidência que sobressai dos dados analisados das avaliações dessas entidades, diz respeito à execução das ações por meio de convênios, cuja sistemática é incompatível com a agilidade e tempestividade que se espera para a execução das ações governamentais viabilizadas por intermédio desse instrumento. Para essa questão, da mesma forma, não há evidências de melhorias ao longo do período examinado, mas, sim, um agravamento do problema com consequências importantes para o alcance dos resultados esperados.

Observou-se, ainda, no que diz respeito à tempestividade do relatório de avaliação para subsidiar a elaboração e a discussão do orçamento do exercício subsequente, bem como a revisão do PPA, que o prazo fixado na Lei para encaminhamento ao Congresso, 15 de setembro, é incompatível com os prazos fixados para elaboração, encaminhamento, discussão e aprovação dos orçamentos relativos ao exercício subsequente à avaliação. Assim, as avaliações referentes ao exercício de 2004, possivelmente serão utilizadas quando da discussão dos orçamentos de 2006, que, por sua vez, terão por base o orçamento de 2005, tornando-as desatualizadas em relação aos resultados a serem considerados.

Além da incompatibilidade dos prazos, o processo legislativo também compromete a tempestividade desejada para aprovação dos orçamentos. Especialmente quanto às revisões do Plano Plurianual 2004-2007. Padilha (2008) constatou que a lógica da revisão das metas físicas do PPA 2004-2007 foi alterada, uma vez que essas metas foram atualizadas pelas Leis Orçamentárias Anuais - LOAs, quando deveria ser o contrário, as revisões do PPA deveriam orientar a elaboração e a definição das metas das LOAs.

Em vista do exposto, pode-se depreender que as avaliações de programas do PPA 2004-2007 não observaram os padrões da utilidade, o que poderia indicar que pouco favorece a utilização tempestiva para cumprir adequadamente os objetivos estabelecidos. No entanto, é possível identificar, com base na análise dos dados primários relativos às respostas do questionário avaliativo disponibilizados pela SPI, que o processo de avaliação tem melhorado a cada ano e a sua utilização vem se refletindo nas mudanças de percepção em relação a sua finalidade, conforme demonstram as respostas a uma questão incluída no questionário aplicado para as avaliações de 2007, na categoria relativa à avaliação do processo.

Nessa categoria analítica foram formuladas as questões apresentadas no QUADRO

Categoria de Análise	Número Questão	Orientação	Número Item
Avaliação do Processo	1	Avalie o processo de avaliação do PPA no período de 2004 a 2007, atribuindo nota de 0 a 10 a cada item que se segue e indicando sua evolução ao longo do período 2004-2007.	1. Capacitação dos gerentes para a avaliação do programa. 2. Orientação do Ministério do Planejamento sobre o processo de avaliação. 3. Clareza do Manual de Avaliação. 4. Conteúdo do Manual de Avaliação. 5. Cronograma do processo de avaliação. 6. Suporte e auxílio da Secretaria de Planejamento e Investimento Estratégicos - SPI/MP às dúvidas quanto à metodologia. 7. Suporte e auxílio da Secretaria de Planejamento e Investimento Estratégicos - SPI/MP às dúvidas no roteiro e problemas de sistema. 8. Melhoria da programação alcançada por meio da avaliação do programa. 9. Utilize o campo abaixo para considerações e sugestões acerca dos aspectos abordados neste item.
	2	Caso existam, descreva práticas adotadas pela gerência do programa, durante os processos de avaliação no período 2004-2007, que poderiam ser incorporadas ao processo de avaliação do PPA.	Resposta (campo não obrigatório para a conclusão do relatório):
	3	Autoavaliação da participação do Gerente de Programa no processo.	1. Avalie sua participação no processo de avaliação do programa. Justifique sua resposta, considerando aspectos tais como motivação, adequação da metodologia e utilidade da avaliação.
	4	Avalie de maneira geral o processo de avaliação do PPA 2004-2007, atribuindo nota de 0 a 10.	1. Avalie de maneira geral o processo de avaliação do PPA 2004-2007, atribuindo nota de 0 a 10.
	5	Utilize o campo abaixo para considerações e sugestões acerca dos aspectos abordados neste bloco.	Resposta (campo não obrigatório para a conclusão do relatório):

QUADRO 31 - QUESTÕES FORMULADAS NA AVALIAÇÃO DOS PROGRAMAS DO PPA 2004-2007 RELATIVAS À AVALIAÇÃO DO PROCESSO
 FONTE: BASE DE DADOS AVALIAÇÃO_PPA_2004-2007 (SPI/MPOG)

No que diz respeito à questão 1, item 8, que indagou sobre a melhoria da programação alcançada por meio da avaliação do programa ao longo do PPA 2004-2007, constatou-se que do total de 161 programas que ofereceram respostas a esta questão, 1 considerou muito pior a cada ano, 3 pior a cada ano, 55 igual em todos os anos, 55 um pouco melhor a cada ano e 47 melhor a cada ano.

Cabe ressaltar que não constaram dos dados disponibilizados pela SPI, as respectivas notas atribuídas, somente o grau de evolução, segundo a escala apresentada nas respostas.

Para as 26 avaliações examinadas, observou-se que, para 11 programas, o grau de evolução foi igual em todos os anos; para 8, foi um pouco melhor a cada ano; para 6 foi melhor a cada ano e 1 considerou que foi muito pior a cada ano.

Com base nesses dados, depreende-se que dos 358 programas vigentes em 2007, 161 deles, ou seja 44%, responderam ao item 8 da questão 1. Desses 161 programas, 102 (63%) consideram que a avaliação contribuiu para a melhoria da programação orçamentária, o que representa 28% do total de programas em execução em 2007. No entanto, para a amostra de programas examinados, 53% consideram que a avaliação contribuiu para a melhoria da programação, fortalecendo as evidências de que as avaliações têm sido utilizadas para o aperfeiçoamento da programação, corroborando a sua utilidade e indicando que, em que pese a não observância do padrão de utilidade, a SPI tem promovido melhorias no instrumento avaliativo, que têm contribuído para o aperfeiçoamento do processo de elaboração orçamentária.

Embora tenha sido incluída, nos Relatórios de Avaliação dos programas do PPA 2004-2007, Ano Base 2007, a análise dessa categoria (avaliação do processo), as informações foram agregadas em sete itens e para o cálculo das médias utilizou-se as médias da pontuação e não do grau de evolução, inviabilizando a comparação com os dados disponibilizados. No entanto, os resultados apresentados pela SPI para as médias-gerais foram: 1) orientações do Ministério do Planejamento sobre o processo de avaliação (7,9); 2) clareza do Manual de Avaliação (8,3); 3) conteúdo do Manual de Avaliação (6,8); 4) cronograma do processo de avaliação (8,4); 5) suporte e auxílio da SPI/MP às dúvidas no roteiro e problemas de sistemas (8,1); e, 7) melhoria da gestão dos programas alcançada por meio da avaliação setorial (7,1).

Com relação aos padrões de viabilidade e propriedade, esses não se aplicam dada a natureza impositiva das avaliações dos programas do PPA.

Para análise dos aspectos metodológicos, abordados nos padrões de acuracidade, cabe inicialmente tecer algumas considerações sobre o modelo único adotado para a avaliação dos programas do PPA 2004-2007.

As avaliações do PPA têm múltiplos objetivos e interessados, que podem ser agrupados em dois grandes grupos. Aqueles destinados a atender aos interesses do público externo (políticos, sociedade, órgãos de controle) e aqueles voltados às necessidades internas da administração pública, em especial aos responsáveis pelo planejamento e pela implementação dos programas (QUADRO 32). No que diz respeito aos stakeholders externos,

as avaliações têm como propósito principal a *accountability*, a verificação da conformidade entre o previsto e o executado, bem como a identificação das falhas e deficiências, conforme definido por Weiss (1998), Patton (2008) e Mark e Henry (2004). No que diz respeito aos *stakeholders* internos, as avaliações têm como principal finalidade o monitoramento, ou seja, contribuir como *feedback* para os executores disponibilizando informações sobre o andamento de determinado programa (Weiss, 1998).

Interessados	Objetivos
Stakeholders Externos	Aferir o resultado, tendo como referência os objetivos e as metas fixadas
	Subsidiar o processo de alocação de recursos públicos, a política de gastos públicos e a coordenação das ações de governo
	Evitar a dispersão e o desperdício de recursos públicos.
Stakeholder Internos	Auxiliar a tomada de decisão
	Promover o aprendizado e a disseminação do conhecimento nas organizações;
	Aperfeiçoar a concepção e gestão do plano
	Promover a <i>accountability</i> das ações do governo.

QUADRO 32 - RELAÇÃO ENTRE OBJETIVOS E INTERESSADOS NAS AVALIAÇÕES DE PROGRAMAS DO PPA

Esse modelo híbrido de avaliação, que atende tanto aos propósitos da *accountability*, como aos do monitoramento, tem sido disseminado por Ray Rist e Jody Zall (2004), sob a denominação *Results-based M&E*, que poderíamos traduzir como sistemas de monitoramento e avaliação baseados em resultados. Eles se diferenciam dos sistemas tradicionais de monitoramento, porque incluem a avaliação dos resultados, buscando integrar o acompanhamento do desempenho com o nível de alcance dos objetivos ou resultados desejados (KUSEK; RIST, 2004).

No entanto, em que pese o modelo de avaliação adotado pelo Governo Federal contemplar uma multiplicidade de objetivos, com vistas a atender diferentes propósitos, a abordagem metodológica utilizada não se mostra adequada e compatível com esses fins.

Conforme mencionamos, a avaliação dos programas do PPA é realizada por intermédio da aplicação anual de um questionário padrão para todos os programas indistintamente, que poderia ser comparado aos modelos de avaliação do tipo *one-size-fits-all* (Radin, 2006), que pela sua própria concepção, dificulta a utilização, considerando que o levantamento de informações inclui aspectos que, muitas vezes, não são adequados ao tipo de programa avaliado ou ao estágio em que se encontra.

Ademais, em que pese terem sido definidos os elementos necessários para a estruturação dos programas constantes do PPA, as deficiências na concepção são tão evidentes, que anualmente são feitos os mesmos questionamentos sobre essa dimensão.

A dinâmica dessas alterações foi ressaltada por Padilha (2008) ao constatar a quantidade de modificações procedidas no PPA aprovado pela Lei nº 10.933/2004, quando da proposta de revisão do Plano de 2004, que expôs “a necessidade de exclusão de 17 programas, a inclusão de 18 e a alteração em 347. A soma dos programas excluídos (17) com os alterados (347) revela que 97,33% dos programas inicialmente previstos sofreram algum tipo de intervenção ainda no primeiro exercício de vigência do PPA.” (PADILHA, 2008).

Além disso, a abordagem utilizada para avaliação dos programas, por meio do questionário, não contempla uma análise prévia da qualidade dos dados e informações dos elementos que compõem o programa: objetivo, público-alvo, justificativa, estratégia de implementação, ações, produtos, metas físicas e financeiras e principalmente dos indicadores adotados, que deveriam ser capazes de medir a evolução do programa. Essa análise é delegada ao responsável pela implementação do programa, que, por outro lado, não tem poder e autoridade para adequar, corrigir ou modificar esses elementos.

Com relação aos indicadores, o TCU, no Relatório sobre as Contas do Governo do exercício de 2004, constatou que apenas 186 programas apresentavam indicadores completos, pouco mais de 50% do total de programas (BRASIL, 2005).

No que diz respeito às metas físicas, Padilha (2008), ao comparar as metas físicas estabelecidas no PPA aprovado para o exercício de 2004, para um conjunto de 1430 ações relativas a 52 programas, constatou que apenas 50,84% desse total manteve as mesmas características ao longo do período de vigência do Plano, sendo que mais de 40% do universo pesquisado constou apenas do PPA relativo ao exercício de 2007. Essa dinâmica reforça a constatação da dificuldade em se avaliar a evolução do desempenho ao longo do período examinado.

A abordagem metodológica aplicada ao questionário não evidencia o contexto do programa, quer seja técnico, organizacional, social, político econômico, de forma a favorecer a análise dos resultados apresentados pelos responsáveis pela implementação das ações.

Assim, não há critérios para mensurar e julgar a adequação, pertinência e suficiência dos recursos, das metas, dos índices ou dos resultados alcançados, em que pese constar dos dados analisados uma estimativa do gerente do programa quanto ao alcance dos resultados e dos índices. Por exemplo, em 2004, das 26 avaliações analisadas, 19 consideraram que os resultados estavam dentro ou acima do esperado. Em 2006, a escala foi alterada, passando a contemplar intervalos percentuais; desta forma, em 2007, desses 19 programas, 15 consideraram que seus resultados estavam entre 80 e 100% do previsto para aquele exercício, e, 4, acima de 100% do esperado.

As evidências demonstram que os relatórios das avaliações do PPA apresentam uma síntese das atividades e ações executadas no âmbito dos programas, demonstrando os resultados alcançados e os recursos despendidos, o que representa um avanço em termos de transparência governamental. No entanto, a abordagem metodológica aplicada não observa os padrões de avaliação, inclusive quanto à acuracidade, ante as características mais informativas do que avaliativas do modelo adotado.

Hatry *et al.* (2004 *apud* Patton, 2008) consideram que os sistemas de monitoramento devem, principalmente, avaliar os resultados do programa sem qualquer investigação mais aprofundada. Mas no caso das avaliações de programas do PPA, não existe um processo avaliativo.

Já os sistemas de monitoramento e avaliação baseados em resultados, segundo Kusek e Rist (2004), são estruturados com base em dois componentes complementares, um de monitoramento e outro de avaliação, cada qual aborda questões específicas, cujas análises e informações terão diferentes formas de utilização. Os achados do monitoramento podem gerar questões a serem respondidas por intermédio de avaliações mais profundas. No entanto, os autores ressaltam que, cada componente, isoladamente, ao final, é insuficiente.

O sistema de monitoramento e avaliação adotado pelo Governo Federal, no entanto, ainda carece de aperfeiçoamento, especialmente quanto à definição dos seus objetivos, da sistemática de monitoramento e avaliação e da abordagem metodológica empregada, de forma a assegurar maior qualidade e credibilidade aos resultados apresentados, com vistas a favorecer e ampliar a sua utilização tanto pelos *stakeholders* internos como externos.

5.2 A UTILIDADE DA AVALIAÇÃO DE PROGRAMAS DO PPA

No questionário de avaliação dos programas do PPA do exercício de 2007, foi incluída uma nova categoria avaliativa, que tratou da avaliação do processo, conforme apresentado na seção anterior, onde também foi inserido um tópico sobre a autoavaliação da participação do Gerente de Programa no processo. A pergunta formulada solicitava ao respondente que indicasse o grau de evolução, segundo a seguinte escala: 1(muito pior a cada ano); 2 (pior a cada ano), 3 (igual em todos os anos); 4 (um pouco melhor a cada ano) e 5 (melhor a cada ano), justificando sua resposta, nos seguintes termos

1. Avalie sua participação no processo de avaliação do programa. Justifique sua resposta, considerando aspectos tais como motivação, adequação da metodologia e utilidade da avaliação (Base de Dados_Avaliação_PPA_2004-2007).

Das respostas analisadas para os 26 programas examinados, constatou-se que para 13 respondentes a sua participação está um pouco melhor a cada ano (grau de evolução 4), para 10, melhor a cada ano (grau de evolução 5) e para 3, igual em todos os anos (grau de evolução 3).

As justificativas apresentadas para esta questão demonstraram a importância da avaliação para a gestão, propiciando um momento de reflexão sobre o programa, seus resultados, deficiências e oportunidades de melhoria, contribuindo para o aprendizado organizacional, permitindo acompanhar com maior clareza a realização e o andamento das atividades necessárias para o desenvolvimento das ações.

Observou-se também, que à medida que o PPA se consolida como uma ferramenta gerencial, os conceitos e práticas de avaliação estão sendo internalizados fatores importantes para o aperfeiçoamento da gestão.

Além disso, a participação tem propiciado o progressivo entendimento do processo e, com a evolução das atividades dos programas, aperfeiçoa-se a compreensão do contexto em que estão inseridas.

Segundo um dos respondentes, “na autoavaliação é possível inferir que há necessidade de maior interação entre os entes envolvidos no processo de adequação da metodologia. É inegável a imensa utilidade desta avaliação para a implementação de melhorias no processo de acompanhamento e gestão das ações.” (BASE de DADOS, AVALIAÇÃO_PPA_2004-2007, SPI/MPOG).

Essas constatações ganham maior relevo na medida em que se observa que as avaliações dos programas do PPA foram as únicas avaliações realizadas em 23, dos 26 programas examinados. Pode-se depreender que, para os programas examinados, a cultura avaliativa está cada vez mais disseminada, podendo-se inferir que há indícios de utilização conceitual e instrumental das avaliações.

No que diz respeito aos 343 programas que ofereceram respostas para a questão 3, observou-se que, para apenas 2 programas, a participação dos respondentes foi considerada muito pior a cada ano (grau de evolução 1), para 3 respondentes foi considerada pior a cada ano (grau de evolução 2), para 66, igual em todos os anos (grau de evolução 3), para 112, um pouco melhor a cada ano (grau de evolução 4) e para 160 respondentes foi considerada melhor a cada ano.

Esses dados revelam que, para quase 80% dos programas que responderam à questão, a participação dos gerentes no processo de avaliação tem sido aperfeiçoada. Já na amostra examinada (26 programas) esse percentual sobe para 88% e quando analisada em

conjunto com as justificativas para o item 8 da questão 1, pode inferir que contribuíram para o aprendizado organizacional e para a melhoria do gerenciamento dos programas. Essa constatação evidencia, ainda, que a participação dos *stakeholders* é o fator determinante para a utilização das avaliações e, em decorrência, para o alcance dos propósitos que deram origem aos trabalhos, conforme demonstra a revisão da literatura apresentada.

6 ANÁLISE DAS AVALIAÇÕES DE PROGRAMAS REALIZADAS PELO TRIBUNAL DE CONTAS DA UNIÃO NO PERÍODO 2004-2007

As avaliações de programas realizadas pelo Tribunal de Contas da União, sob a denominação de auditorias de natureza operacional – ANOp, –, têm por objetivo examinar a eficiência, a eficácia, a efetividade e a economicidade dos programas governamentais, com vistas a contribuir para o aperfeiçoamento da administração pública, por intermédio da formulação de recomendações para promoção de melhorias na gestão e nos programas.

Essas avaliações atendem ao propósito da *accountability*, na medida em que objetivam examinar se os programas estão sendo implementados adequadamente, se as metas e os objetivos estão sendo atingidos, se os recursos estão sendo empregados de forma eficiente, se os problemas estão sendo controlados e tratados e se os resultados são efetivos (PATTON, 2008).

Mark, Henry e Julnes (2000) classificam os propósitos desse tipo de avaliação como de conformidade e identificação de falhas, na medida em que buscam estimar em que extensão o programa está alcançando determinadas expectativas, como, por exemplo, as diretrizes estatutárias, os regulamentos ou outras exigências, inclusive os níveis de desempenho especificados.

As avaliações de conformidade e identificação de falhas tendem a examinar a fidelidade com que o programa é implementado. A Fidelidade, nesse caso, é geralmente entendida como aderência a mandatos específicos do programa e procedimentos estabelecidos externamente, como as práticas geralmente aceitas de contabilidade ou normas que estabelecem igualdade de oportunidade nas decisões sobre pessoal. Desta forma, esse propósito em geral é uma importante função para o Legislativo, Tribunais e altos níveis da administração, que, em tese, devem assegurar que as políticas sejam executadas adequadamente e em complemento, que os resultados estão nos níveis esperados (MARK; HENRY; JULNES, 2000).

As avaliações de programas realizadas pelo TCU foram impulsionadas com as mudanças introduzidas pelo Plano Diretor de Reforma do Estado de 1995, cujas alterações no desenho e nas atribuições do Estado passaram a exigir uma reestruturação dos processos internos de atuação do controle a cargo do Tribunal de Contas da União (SERPA, 2004).

Essas mudanças foram objeto de análise interna pelo TCU, por intermédio de estudo realizado pelo Grupo de Pesquisa Tendências do Controle da Gestão Pública, (Brasil, 1999), que constatou que as alterações promovidas no âmbito da Administração Pública Federal tinham reflexos nas atividades desenvolvidas pelo Tribunal, podendo ser estruturadas em dois segmentos. O primeiro trata da reorganização do modo de funcionamento da máquina

governamental, com foco nos resultados da ação do poder público e envolve a mudança no paradigma de *accountability*, o contratualismo, a ampliação da utilização da tecnologia da informação, a reformulação do processo orçamentário e do controle interno. O segundo diz respeito à redução do tamanho do Estado com maior presença da sociedade nas atividades governamentais.

Em decorrência dos resultados apresentados pelo Grupo de Pesquisa, promoveu-se a reestruturação organizacional, com a criação de secretarias especializadas (2000), buscando-se uma atuação matricial. Assim, foram mantidas as secretarias temáticas na sede e nos estados, organizadas de acordo com as funções dos órgãos jurisdicionados ao TCU e sua localização (Saúde, Educação, Trabalho, Justiça, Transportes, etc) e criadas as secretarias especializadas, que passaram a atuar em temas específicos, como avaliação de programas (Secretaria de Fiscalização e Avaliação de Programas - Seprog), fiscalização de obras (Secretaria de Fiscalização de Obras e Patrimônio da União - Secob), fiscalização da desestatização e controle da regulação (Secretaria de Fiscalização de Desestatização - Sefid), fiscalização de pessoal (Secretaria de Fiscalização de Pessoal - Sefip), macroavaliação governamental (Secretaria de Macroavaliação Governamental – Semag) e, de recursos (Secretaria de Recursos – Serur). Em 2006 foi criada a Secretaria de Fiscalização de Tecnologia da Informação (Sefti).

Para o exercício das atividades dessas novas secretarias, os processos de trabalho foram aperfeiçoados, plasmando-se as necessidades do Controle no novo contexto da administração pública, com vistas a assegurar a efetiva e regular gestão dos recursos públicos em benefício da sociedade.

Desta forma, no que diz respeito à avaliação de programas, o Tribunal, seguindo o movimento internacional dos órgãos de controle, – conforme descrito por Barzelay (1997) e com base nas diretrizes estabelecidas pela Organização Internacional das Entidades de Fiscalização Superior (Intosai, 2003) e o apoio do Departamento para o Desenvolvimento Internacional do Reino Unido (DFID) – buscou o desenvolvimento metodológico adequado à prática avaliativa, por intermédio de dois projetos complementares.

O primeiro projeto, implementado entre 1998 e 2001, denominado Projeto Desenvolvimento de Técnicas de Auditoria Operacional, teve por objetivo o desenvolvimento da metodologia de auditoria operacional. O Segundo, intitulado Projeto Aperfeiçoamento do Controle Externo com Foco na Redução da Desigualdade Social (CERDS), vigente no período de 2002 a 2006, objetivou ampliar o escopo da metodologia desenvolvida para as

auditorias operacionais, por intermédio da inclusão de técnicas de avaliação focadas na redução da pobreza, na equidade, no fortalecimento do controle social e na aproximação com a sociedade civil (DFID, 2006).

De fato, o projeto buscou contribuir para uma maior transparência na prestação de contas do gasto público (*accountability*) ao avaliar programas governamentais e modelos regulatórios de serviços públicos com foco na redução da desigualdade social. (DFID, 2006).

6.1 ANÁLISE DA ADEQUAÇÃO DAS AVALIAÇÕES DE PROGRAMAS REALIZADAS PELO TCU AOS PADRÕES DE AVALIAÇÃO DE PROGRAMAS DO COMITÊ CONJUNTO DE PADRÕES PARA AVALIAÇÃO EDUCACIONAL DOS ESTADOS UNIDOS

A análise dos dados coletados dos questionários demonstrou que as avaliações realizadas pelo TCU indicam aderência significativa aos padrões de avaliação, para aqueles que lhes são aplicáveis. Foram excluídos desta análise os padrões de propriedade relativos à contratação (P2), bem como o (P8) que trata da questão da responsabilidade fiscal (prestação de contas dos recursos alocados e das despesas realizadas) e um padrão de acuracidade, (A12) meta-avaliação, uma vez que as avaliações examinadas não foram reavaliadas.

No que diz respeito aos padrões da utilidade, aqueles que estruturam e orientam as avaliações, definindo de forma prática se a avaliação atende às necessidades de informações daqueles cujos problemas estão sendo investigados, contemplando a identificação dos *stakeholders* (U1); a análise da credibilidade do avaliador (U2); as informações sobre a seleção e o escopo do trabalho (U3); a identificação dos valores (critérios) (U4); a objetividade do relatório (U5), a tempestividade e disseminação do relatório (U6); bem como, a avaliação do impacto; as médias atribuídas a cada padrão indicam que a estratégia metodológica adotada nas avaliações do TCU preza pela identificação dos atores envolvidos no processo avaliativo, pela competência dos profissionais avaliadores, pela identificação e seleção das informações pertinentes para elaboração das questões a serem investigadas e que mais afetam o desempenho do programa, pela clareza do relatório, bem como pelo envolvimento dos principais atores que devem acompanhar a avaliação, com vistas a sua utilização (TABELA 2).

TABELA 2 - MÉDIAS DOS PADRÕES DE UTILIDADE POR AVALIAÇÃO

Utilidade	Aquicultura (Sup)	Aquicultura (Coord)	Medicamentos (Sup)	Medicamentos (Coord)	Transplantes	Hidroviás	Exploração	Geologia	Terras Indígenas	Pesquisadora	Média
U1	3,57	3,57	3,29	3,29	3,00	2,86	2,71	3,43	3,29	3,29	3,23
U2	3,33	3,33	3,50	3,50	3,00	3,75	2,50	3,50	2,50	3,05	3,20
U3	4,00	4,00	3,67	3,67	2,33	4,00	2,33	3,00	3,67	3,67	3,43
U4	4,00	4,00	4,00	4,00	4,00	3,00	3,00	4,00	3,00	3,67	3,67
U5	2,50	2,50	3,50	3,50	4,00	4,00	2,75	3,33	3,00	3,83	3,29
U6	3,50	3,50	2,00	2,00	2,33	2,00	1,67	2,67	2,33	2,33	2,43
U7	3,00	3,00	3,33	3,33	2,67	3,00	3,00	3,33	3,33	3,22	3,12

FONTE: A autora (2010)

No que diz respeito à tempestividade e disseminação do relatório, cuja média geral ficou um pouco acima de 2, indicando que esse padrão é observado com restrições que limitam o pleno atendimento da questão examinada, observou-se que esse padrão (U6), segundo Sanders (1994) é analisado considerando, principalmente, os seguintes aspectos:

- i) se os proponentes da avaliação e os representantes dos grupos-chaves de *stakeholders* deixaram claro seus principais objetivos com relação à avaliação, a forma de apresentação do relatório e a data mais apropriada para sua apresentação;
- ii) se houve alguma negociação quanto ao plano de disseminação da avaliação para os interessados na sua utilização e quanto ao acesso deles aos achados e recomendações específicas; e
- iii) se a avaliação foi submetida a algum controle de qualidade e seu resultado foi apresentado aos interessados na utilização.

Quanto a esses aspectos, cabem algumas considerações que diferenciam as avaliações realizadas pelos órgãos de controle.

As avaliações realizadas pelo Tribunal, sob a denominação de auditoria operacional, são propostas por iniciativa do próprio órgão, no exercício do seu mandato constitucional, ou por solicitação do Congresso Nacional. Assim, os objetivos a serem atingidos com a realização da avaliação, o formato do relatório e a data para sua apresentação são estabelecidos pelo TCU, em que pese a abordagem metodológica adotada nos trabalhos examinados buscar a participação dos principais *stakeholders* responsáveis pela implementação dos programas, desde a etapa de seleção dos programas, conforme se depreende nos estudos de viabilidade examinados, mediante consulta a esses atores quanto ao seu interesse e oportunidade para a realização da avaliação.

Quanto ao controle de qualidade, observou-se que os trabalhos foram executados segundo um documento denominado Controle de Qualidade, que estabelece procedimentos e mecanismos de controle com vistas a assegurar a qualidade e tempestividade dos trabalhos, mas nem todos os resultados desse controle foram apresentados aos interessados na utilização. Em geral, constatou-se que as matrizes de planejamento e de achados, após discutidas nos painéis de referência, foram levadas ao conhecimento dos responsáveis pelos programas avaliados.

No GRÁFICO 1, pode-se visualizar a análise dos padrões para cada avaliação, bem como a média geral (linha cheia) e as médias da revisão dos trabalhos realizada por esta pesquisadora (linha pontilhada). Pode-se constatar que há uma convergência para o intervalo onde os padrões foram considerados adequados (entre 3 e 4), sendo o padrão U6 o menos observado, em razão das características acima mencionadas.

Sobressai da análise a baixa aderência aos padrões de utilidade da avaliação realizada no Programa Combate ao Abuso e Exploração de Crianças e Adolescentes. No entanto, no relatório de avaliação não constam informações que permitam identificar os fatores que comprometeram a observância dos padrões, nesta avaliação, em relação aos demais trabalhos.

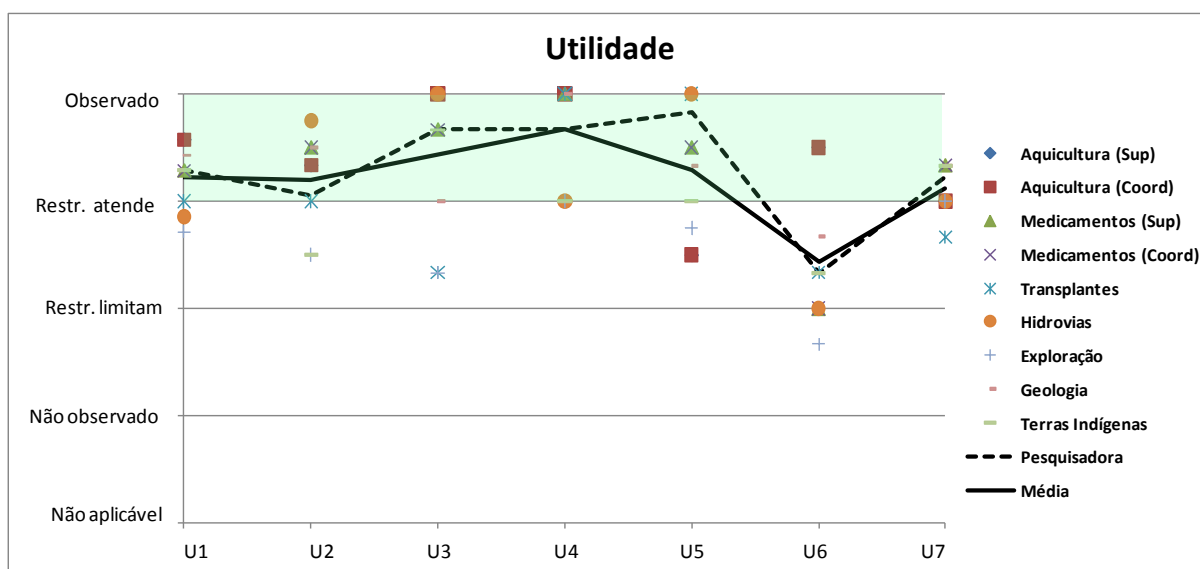


GRÁFICO 1 - MÉDIAS DOS PADRÕES DE UTILIDADE POR AVALIAÇÃO

FONTE: A autora (2010)

Os padrões de viabilidade são aqueles que têm por objetivo assegurar que as avaliações serão realistas, prudentes, cordiais e econômicas. Espera-se que sejam concebidas considerando o ambiente no qual o programa é executado e o contexto real do objeto avaliado. Na execução dos trabalhos não devem utilizar mais recursos financeiros, pessoas, materiais ou tempo do que o necessário para responder as questões estudadas.

Quanto aos três padrões de viabilidade examinados: procedimentos práticos (F1), viabilidade política (F2) e custo-efetividade (F3), a análise evidencia que esses padrões são adequadamente observados, exceto quanto ao custo-efetividade (TABELA 3).

TABELA 3 - MÉDIAS DOS PADRÕES DE VIABILIDADE POR AVALIAÇÃO

Viabilidade	Aquicultura (Sup)	Aquicultura (Coord)	Medicamentos (Sup)	Medicamentos (Coord)	Transplantes	Hidrovias	Exploração	Geologia	Terras Indígenas	Pesquisadora	Média
F1	4,00	4,00	4,00	4,00	3,75	4,00	4,00	4,00	3,00	3,75	3,85
F2		4,00	4,00	4,00	3,00	3,00	3,00	4,00	4,00	3,00	3,56
F3	4,00	3,00	4,00	4,00	2,00	2,00	2,00	3,00	2,00	2,33	2,83

FONTE: A autora (2010)

O padrão (F3) tem por objetivo assegurar que a relação entre os custos da avaliação e os resultados apresentados sejam adequados, como também que as informações disponibilizadas tenham suficiente valor, de forma a evitar, utilizando uma expressão coletada durante as entrevistas realizadas, que se “utilize uma bala de canhão para matar uma formiga”, no sentido de se requerer uma avaliação altamente sofisticada para tratar questões pouco complexas, cujos mesmos resultados poderiam ser obtidos a um custo menor.

Nas avaliações examinadas, constatou-se que são apresentados, nos relatórios de planejamento, uma planilha indicando os custos diretos da auditoria e os possíveis benefícios esperados com a realização do trabalho. No entanto, nesta etapa, não há quantificação dos benefícios, tampouco se evidenciou qualquer análise quanto à adequação dos custos em relação aos resultados esperados.

O GRÁFICO 2 apresenta as médias aferidas para cada uma das avaliações pesquisadas, no que diz respeito aos padrões de viabilidade.

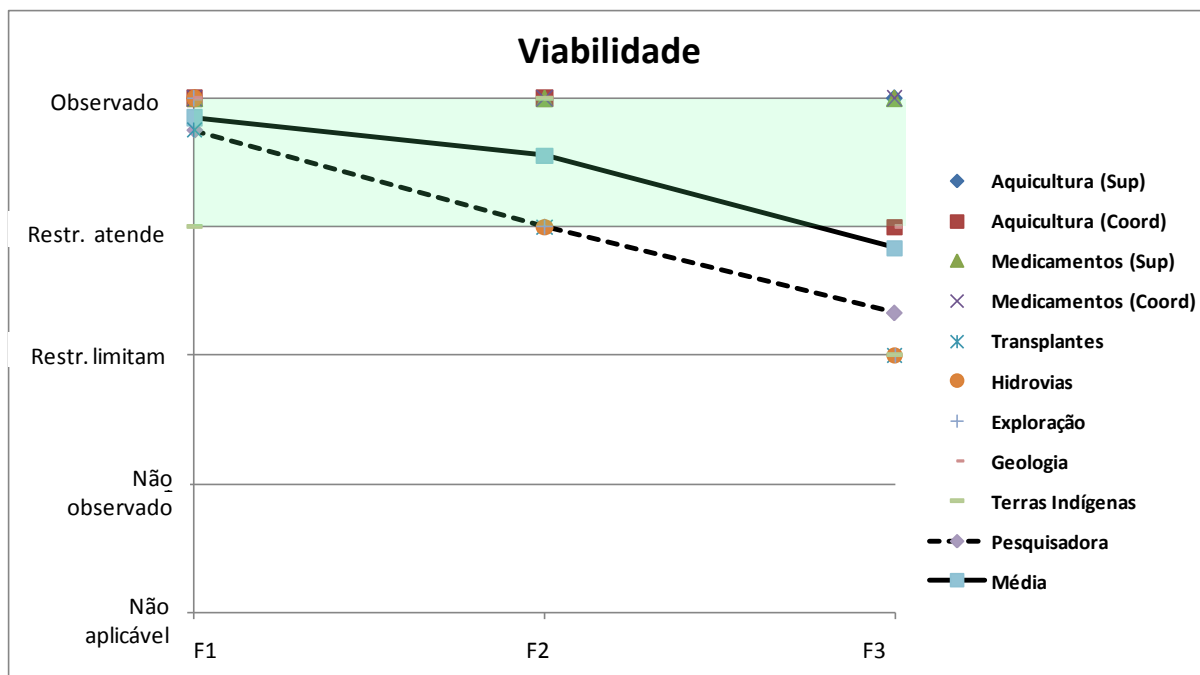


GRÁFICO 2 - MÉDIAS DOS PADRÕES DE VIABILIDADE POR AVALIAÇÃO

FONTE: A autora (2010)

Os padrões de propriedade têm por objetivo assegurar que a avaliação será conduzida com observância dos princípios legais e éticos, respeitando os direitos individuais e o bem-estar das pessoas envolvidas no processo, como também daquelas que podem vir a ser afetadas pelos resultados dos trabalhos. Esses padrões são examinados a partir da verificação de questões que abordam, por exemplo: (P1) se o foco da avaliação foi nas características do programa que mais afetam os beneficiários e que promovem o alcance dos objetivos da organização ou do programa; (P2) se a contratação da avaliação foi devidamente formalizada e documentada; (P3) se as avaliações foram concebidas e conduzidas respeitando e protegendo os direitos e o bem-estar dos indivíduos; (P4) se nas relações entre os avaliadores e as pessoas que participaram da avaliação foram respeitados os valores dos indivíduos; (P5) se a avaliação foi completa e imparcial no exame das forças e fraquezas que incidem sobre o programa examinado; (P6) se a totalidade dos achados e as suas respectivas limitações estavam acessíveis para as pessoas afetadas pela avaliação, bem como para aqueles que deveriam receber os seus resultados; (P7) se há conflitos de interesses que podem comprometer o processo e os resultados da avaliação; e, (P8), se os recursos alocados e as despesas realizadas pelo avaliador estavam devidamente refletidos nos procedimentos de prestação de contas, possibilitando a sua comprovação e registro. O resultado da análise dos dados e da documentação indica que esses padrões foram observados (TABELA 4), com algumas restrições que apontamos a seguir.

TABELA 4 - MÉDIAS DOS PADRÕES DE PROPRIEDADE POR AVALIAÇÃO

Propriedade	Aquicultura (Sup)	Aquicultura (Coord)	Medicamentos (Sup)	Medicamentos (Coord)	Transplantes	Hidroviás	Exploração	Geologia	Terras Indígenas	Pesquisadora	Média
P1	3,33	3,67	3,67	3,67	2,67	3,33	3,67	3,67	3,33	2,67	3,37
P3	2,67	3,00	2,67	2,67		2,33	2,33		3,00	2,86	2,69
P4	3,33	3,67	4,00	4,00	4,00	3,67	2,67	3,67	2,33	3,89	3,52
P5	3,50	3,67	3,67	3,67	3,33	2,33	3,67	3,67	3,00	3,22	3,37
P6	3,00	3,00	4,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00	3,70
P7	1,00	2,50	2,50	2,50	4,00	1	2,50	3,00	2,50	1,50	2,30

FONTE: A autora (2010)

No que diz respeito aos padrões (P2) e (P8), cabe observar que as avaliações do TCU, conforme mencionado anteriormente, são realizadas por iniciativa própria ou por solicitação do Congresso Nacional e são executadas por servidores do quadro de pessoal do Tribunal, com a possibilidade de contratação de consultores externos e a participação de servidores de outras organizações governamentais. O TCU não utiliza a contratação de serviços de avaliação para dar cumprimento a suas atribuições constitucionais, em decorrência disso, não há o pagamento por esses serviços. No entanto, as despesas incorridas durante os trabalhos de avaliação são comprovadas em processo específico, segundo as normas internas do TCU. Ante essas considerações, os padrões (P2) e (P8) não se aplicam às avaliações realizadas pelo Tribunal.

Já com relação ao padrão (P1), que se refere ao foco dos trabalhos de avaliação, as médias apuradas e a análise da documentação evidenciam que as avaliações foram orientadas para as características que mais afetam os beneficiários dos programas e que são críticas para o alcance dos resultados esperados e para o cumprimento dos objetivos. No entanto, outro fator impacta na avaliação de (P1), trata da periodicidade com que são comunicados os resultados intermediários aos atores responsáveis pela implementação de medidas corretivas. De um modo geral, a apresentação das matrizes de planejamento e de achados para os responsáveis, antes da elaboração dos relatórios, foi considerada uma forma de levar ao conhecimento desses atores os resultados intermediários dos trabalhos de auditoria. No entanto, na avaliação do Programa Doação e Transplante de Órgãos e Tecidos, esse item foi considerado não observado pelo coordenador, diminuindo a média de (P1), muito embora tenha sido confirmado, posteriormente, pelo supervisor dos trabalhos de avaliação, que a referida documentação foi apresentada aos responsáveis durante a execução dos trabalhos, em cumprimento aos procedimentos estabelecidos para o processo de avaliação de programas.

Com relação à proteção dos direitos individuais (P3), verificou-se que, quando da realização das pesquisas por intermédio do envio de questionários, os expedientes que encaminharam o instrumento de coleta de dados, ao apresentar os objetivos do trabalho e da

pesquisa, além de enfatizar a importância da participação dos atores selecionados, ressaltaram expressamente que as informações obtidas seriam utilizadas exclusivamente para fins da avaliação, divulgando-se, apenas, os dados consolidados. Esse procedimento é uma forma de se preservar a identidade dos informantes e a confidencialidade dos dados. No entanto, não consta nenhum documento que assegure que os direitos dos participantes estejam protegidos, conforme atestam todas as respostas recebidas para este item. Por esta razão, muito embora a cultura e os valores dos participantes tenham sido compreendidos pelos avaliadores, bem como que tenha sido formalizado o compromisso de se manter a confidencialidade dos dados coletados, a ausência de normatização sobre a coleta e tratamento das informações, que envolvam garantias de direitos individuais, é uma fragilidade, que pode vir a comprometer a credibilidade de avaliações.

Na análise do padrão (P4), que trata das relações pessoais entre os avaliadores e os demais indivíduos envolvidos nos trabalhos, foi verificado em que medida os relatórios e os demais documentos examinados evidenciavam que a cultura, os valores sociais e as diferenças de linguagem foram devidamente compreendidos pelos avaliadores; como também, quais foram os canais de comunicação utilizados para acessar os participantes; e qual o nível de conhecimento dos avaliadores sobre as organizações onde os trabalhos foram realizados.

O padrão (P4) está diretamente relacionado com os padrões da (U4), identificação dos valores, (P1), foco dos trabalhos e (P3) garantia e proteção dos direitos individuais, cujas evidências, conforme analisadas, indicam que foram observados, mas com restrições em relação a (P3), em função da ausência de procedimentos formais que assegurem a proteção dos direitos individuais quando da coleta de dados sobre questões que possam vir a comprometer a garantia desses direitos.

A análise das médias aferidas para (P4) e da documentação examinada, indicam que esse padrão foi observado. No entanto, duas avaliações apresentaram médias menores que 3: a Avaliação do Programa Combate a Exploração Sexual de Criança e Adolescentes (2,67) e, a Avaliação das Ações de Fiscalização e de Acompanhamento de Ingressos em Terras Indígenas (2,33), principalmente em decorrência da ausência ou deficiência de canais de comunicação com participantes dos trabalhos.

No que diz respeito ao padrão (P5), que trata do relatório de avaliação, quanto a sua consistência em relação aos aspectos de isenção e imparcialidade com que foram abordadas as forças e fraquezas que afetam o programa, bem como quanto às limitações que podem ter comprometido o levantamento dos dados, observou-se que a média geral das pontuações

indica que esse padrão foi observado. No entanto, a média aferida para o Programa Manutenção de Hidrovias foi menor do que 3, em razão, principalmente, da não observância às limitações que podem ter comprometido o levantamento dos dados.

A ausência da análise do impacto das limitações sobre os resultados foi uma fragilidade identificada em todos os relatórios examinados. Em geral, foram informadas as limitações relativas à estratégia metodológica adotada, principalmente quanto a não obtenção, insuficiência ou baixa qualidade dos dados e baixo retorno dos questionários.

Com relação à apresentação dos achados, padrão (P6), constatou-se que os relatórios examinados apresentaram os achados identificados e sintetizados nas matrizes de achados, que são compatíveis com os propósitos da avaliação e os métodos utilizados. Como também as conclusões e recomendações são coerentes com as análises procedidas, as quais foram evidenciadas no relatório de forma adequada.

O Padrão (P7) diz respeito à verificação da existência de conflitos de interesses, que podem ocorrer tanto quando o avaliador tem interesses pessoais ou financeiros que podem influenciar ou serem afetados pela avaliação, como quando os *stakeholders* envolvidos têm interesses pessoais ou financeiros na avaliação ou no programa que está sendo avaliado. Desta forma, a questão a ser observada, frequentemente, diz respeito a como lidar com esse problema e não como evitá-lo (SANDERS, 1994).

Esse é um problema recorrente nas avaliações internas, onde as relações de amizade entre os colegas de trabalho favorecem e podem influenciar os resultados das avaliações. É também um problema frequente nas avaliações externas, porque os clientes têm muita liberdade para escolher os avaliadores externos e esses podem conseguir os contratos desde que estejam no negócio. Os conflitos de interesse podem comprometer as avaliações, por meio da corrupção dos processos, dos achados e das interpretações (SANDERS, 1994, p. 116, tradução nossa).

Para examinar essa questão, buscaram-se evidências nos relatórios de avaliação examinados, quanto à identificação de possíveis conflitos de interesses. Além disso, no questionário aplicado na pesquisa foi solicitado aos respondentes que indicassem se, nas avaliações realizadas, foi possível identificar vantagens (financeiras, social, moral ou política) que as partes envolvidas na avaliação poderiam ganhar ou perder com os resultados do trabalho. Como resultado, foram identificadas quatro avaliações em que esse problema foi observado (nas avaliações dos programas/ações: Desenvolvimento da Aquicultura, Medicamentos Excepcionais, Combate ao Abuso e Exploração Sexual de Crianças e Adolescentes, e Fiscalização de Terras Indígenas). Na avaliação do programa Desenvolvimento da Aquicultura, os conflitos estavam relacionados à implementação da

política, quanto à definição das áreas para a execução de projetos com interesses comerciais e sociais; quanto à avaliação das ações de distribuição de medicamentos excepcionais, aos interesses comerciais dos laboratórios privados, que poderiam estar associados à crescente demanda judicial para compra de medicamentos não previstos nos protocolos; no Programa Combate ao Abuso e Exploração de Crianças e Adolescentes, os conflitos de ordem moral em relação às crianças vitimadas e seus violentadores; no Programa de Fiscalização de Terras Indígenas, os interesses políticos, ideológicos, sociais e organizacionais dificultavam a implementação das ações do programa.

No entanto, quando perguntado aos respondentes sobre a existência de procedimentos que protegessem contra os problemas associados aos conflitos de interesses, apenas a avaliação do Programa Doação, Captação e Transporte de Órgãos e Tecidos confirmou a existência, 5 respondentes apontaram que não foi observado, 1 que não se aplicava e outro não se manifestou.

Os problemas associados aos conflitos de interesse merecem especial atenção no desenvolvimento das avaliações, principalmente, nos trabalhos realizados pelos órgãos de controle, em razão do propósito para o qual se destinam: promover a *accountability*. O resultado e as conclusões das avaliações desses órgãos revestem-se da legitimidade delegada constitucionalmente para defesa e salvaguarda do interesse público. Desta forma, três requisitos devem ser observados na condução desses trabalhos: aqueles que dizem respeito à independência do avaliador; a credibilidade da avaliação, cujo principal componente é a qualidade dos trabalhos; e, o terceiro trata da reputação do avaliador, os aspectos morais (CHELIMSKY, 2009).

No GRÁFICO 3, podem-se visualizar as médias aferidas na pesquisa realizada sobre a observância dos padrões de avaliação.

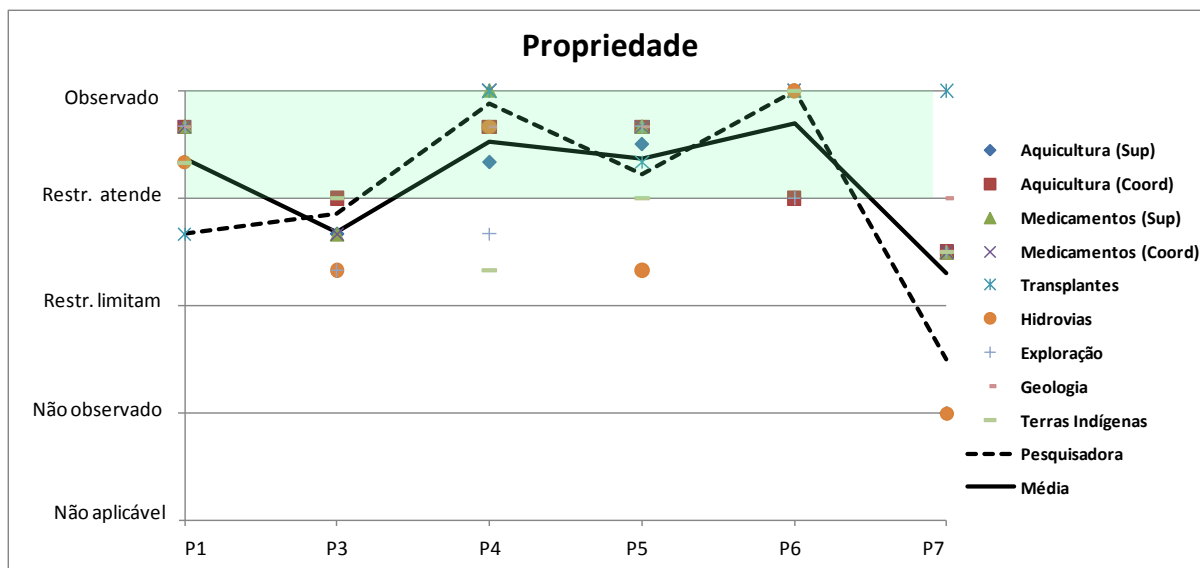


GRÁFICO 3 - MÉDIAS DOS PADRÕES DE PROPRIEDADE POR AVALIAÇÃO
 FONTE: A autora (2010)

Os padrões de acuracidade têm por objetivo assegurar que as informações disponibilizadas pelas avaliações, sobre as características que determinam o valor ou mérito do programa examinado, sejam tecnicamente adequadas e confiáveis.

A acuracidade foi analisada considerando os seguintes aspectos: se o programa foi adequadamente descrito e se a documentação utilizada está disponível para consulta (A1); se o contexto em que o programa está inserido foi examinado com vistas a identificar a sua influência sobre o programa avaliado (A2); se os propósitos e os procedimentos utilizados na avaliação foram detalhadamente descritos e revisados, de forma a possibilitar a sua identificação e avaliação (A3); a credibilidade, validade e confiabilidade das fontes de informações utilizadas (A4), (A5) e (A6); se foram adotados mecanismos de controle para a identificação de erros e inconsistências na coleta, processamento e análise dos dados e informações (A7); a consistência e adequação dos métodos, das análises e das conclusões (A8), (A9) e (A10); e, se o relatório é imparcial (A11).

Para o exame da acuracidade também é recomendado verificar se a própria avaliação foi avaliada formativa e somativamente em relação aos padrões pertinentes, como forma de garantir que foi adequadamente conduzida e orientada, o que seria uma meta-avaliação (A12). No entanto, as avaliações realizadas pelo TCU não foram submetidas à reavaliação, assim o padrão (A12) não foi aplicado.

Para análise dos padrões de acuracidade foram consideradas as médias aferidas nas respostas ao questionário para cada trabalho, apresentadas na TABELA 5, bem como a revisão dos respectivos relatórios de avaliação e o exame da documentação armazenada em

meio magnético e disponibilizada para pesquisa, relativa às avaliações dos programas/ações: Medicamentos Excepcionais, Transplantes de Órgãos e Tecidos e Manutenção de Hidrovias.

TABELA 5 - MÉDIAS DOS PADRÕES DE ACURACIDADE POR AVALIAÇÃO

Acuracidade	Aquicultura (Sup)	Aquicultura (Coord)	Medicamentos (Sup)	Medicamentos (Coord)	Transplantes	Hidrovias	Exploração	Geologia	Terras Indígenas	Pesquisadora	Média
A1	4,00	4,00	3,50	3,50	4,00	4,00	3,00	4,00	4,00	3,83	3,78
A2	3,00	3,50	3,50	3,50	3,00	3,00	3,00	4,00	4,00	3,17	3,37
A3	2,67	2,67	2,67	2,67	3,00	4,00	2,67	3,33	2,67	2,89	2,92
A4	2,50	4,00	3,29	3,29	3,17	4,00	3,67	3,75	2,71	3,19	3,36
A5	3,50	3,67	3,33	3,33	3,33	3,67	2,33	3,33	2,67	3,67	3,28
A6	4,00	4,00	4,00	4,00	4,00	4,00	3,00	2,00	4,00	4,00	3,70
A7	2,00	3,33	2,33	2,33	3,33	2,33	1,67	2,67	2,33	2,56	2,49
A8		3,00	3,67	3,67	4,00	4,00	2,67	3,67	2,33	3,11	3,35
A9	3,50	3,33	3,33	3,33	4,00	4,00	2,00	3,33	3,67	3,33	3,38
A10	3,75	3,75	4,00	4,00	3,75	3,75	3,50	3,75	4,00	3,83	3,81
A11	3,00	4,00	4,00	4,00	3,00	3,00	3,00	2,00	3,00	3,00	3,20

FONTE: A autora (2010)

Pode-se observar, na Tabela 5, que as médias aferidas na pesquisa indicam que os padrões de acuracidade são em grande parte observados, mas apresentam algumas deficiências em relação aos padrões (A3), descrição dos propósitos e procedimentos, e (A7), sistematização das informações.

No que diz respeito ao padrão (A3), constatou-se que os propósitos e as intenções de utilização dos achados foram inicialmente definidos, como recomendado; bem como, que os papéis de trabalho e os procedimentos utilizados para coleta e análise dos dados e informações estão disponíveis para revisão e reanálise. No entanto, esse processo não foi monitorado por um avaliador independente, fato que refletiu desfavoravelmente nos resultados das médias analisadas.

Quanto ao padrão (A7), objetivou-se examinar se existem mecanismos sistematizados de controle que permitam identificar e corrigir erros que possam ocorrer na coleta, no processamento e na análise das informações apresentadas no relatório.

Controle sistemático de informações, segundo Sanders (1994), é um processo a ser observado, cujas etapas a serem seguidas objetivam assegurar que todo dado e informação, já existentes ou coletados, estão, dentro do possível, livres de erros e mantidos em segurança.

Para tanto, esse padrão foi examinado por intermédio de questões que buscaram identificar a existência de alguma sistemática de controle, cujos mecanismos possibilitassem checar os erros na coleta, no processamento e nos resultados das análises; assim como checar se existiam mecanismos de controle para assegurar a integridade das informações originais e se existiam procedimentos para guarda e manutenção das informações coletadas.

As respostas ao questionário indicam que existe uma sistemática de controle para identificar erros na coleta, no processamento e na análise. No entanto, exceto para as Avaliações realizadas nos Programas de Desenvolvimento da Aquicultura e Transplantes, essa sistemática não foi suficientemente adequada para cumprir com o objetivo pretendido.

Os dados e informações coletados em meio magnético são armazenados por avaliação, em pastas da Seprog, localizadas no servidor do TCU, organizadas de acordo com a etapa dos trabalhos (seleção, planejamento, execução, relatório, monitoramento). As informações coletadas por meio das entrevistas, grupos focais, observação direta, ou mesmo os questionários aplicados em papel são guardados na Secretaria. Não foram localizadas nas pastas eletrônicas as transcrições das entrevistas, grupos focais e os relatórios das observações para as avaliações examinadas.

Com relação à organização das pastas eletrônicas, observou-se que muito embora os documentos e demais arquivos de dados e documentos estejam identificados, não há mecanismos de controle que permitam verificar se os dados e informações estão completos, seguros e íntegros.

Pode-se depreender que os procedimentos para indexação, guarda e manutenção da documentação existem, mas necessitam ser aperfeiçoados para assegurar a integridade e segurança das informações utilizadas nas avaliações.

Quanto à sistemática de controle que possibilite checar os erros na coleta, no processamento e nos resultados das análises, observa-se que essa questão tem relação direta com o padrão (P5), no que diz respeito à ausência da análise do impacto das limitações sobre os resultados apresentados nos relatórios, que é uma fragilidade que pode comprometer a credibilidade do trabalho apresentado.

Da leitura dos relatórios, inclusive dos relatórios de planejamento das avaliações examinadas, não foi possível identificar quais os mecanismos utilizados para checar os erros na coleta e no processamento dos dados. Além disso, a leitura da seção relativa à metodologia não indica a forma de organização dos dados, sua localização, limitações, mecanismos de controle, que permita uma reanálise dos dados coletados.

No entanto, cabe ressaltar que as análises são fundamentadas nos dados coletados, que se mostraram consistentes com as conclusões apresentadas para as questões examinadas.

No GRÁFICO 4, é possível visualizar que os padrões de acuracidade foram observados em grande parte dos trabalhos examinados, com as ressalvas mencionadas.

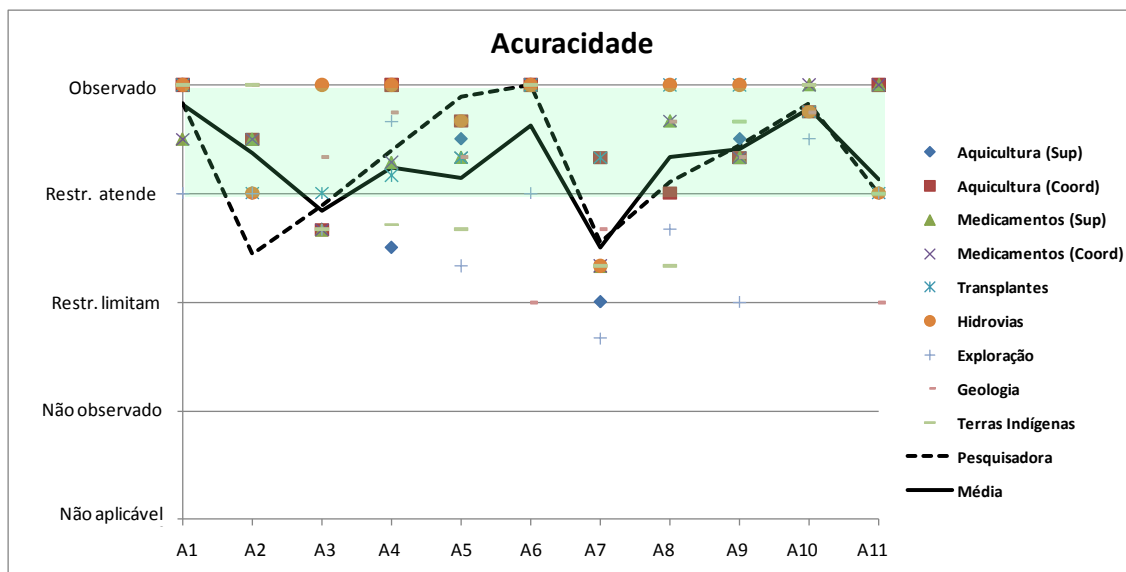


GRÁFICO 4 - MÉDIAS DOS PADRÕES DE ACURACIDADE POR AVALIAÇÃO

FONTE: A autora (2010)

Pode-se depreender da análise da documentação disponibilizada para esta pesquisa que a abordagem metodológica utilizada pelo TCU, no período analisado, foi estruturada a partir da identificação de problemas que comprometem o desempenho e os resultados esperados para os programas e/ou ações.

Os problemas são contextualizados por meio do levantamento de informações sobre o programa, que incluem a identificação dos principais *stakeholders* envolvidos ou afetados pelo programa, por intermédio da análise *stakeholder*, onde são identificados os papéis e interesse de cada ator no programa, o impacto da sua participação para o alcance dos objetivos, com base nos quais, é definido o grau de prioridade da sua participação para definição das questões a serem investigadas (o escopo do trabalho).

A partir da identificação dos principais *stakeholders*, buscou-se, por intermédio da aplicação de técnicas de planejamento estratégico (análise SWOT, diagrama de verificação de risco, de mapa de processo e mapa de produto), identificar na perspectiva dos principais atores, as áreas de maior risco e vulnerabilidade para o alcance dos objetivos do programa.

Com base nessas informações foram elaborados os problemas e as questões a serem investigadas, para as quais foram desenvolvidos, pela própria equipe de avaliação, os instrumentos de coleta de dados.

Uma característica que se observou nas avaliações examinadas do TCU foi o interesse em se obter informações, direta e pessoalmente, dos beneficiários dos programas, quando possível. Essa característica reforça a necessidade de se normatizar tanto o processo como os

procedimentos para coleta de dados quando são investigadas, nas auditorias, questões que envolvem beneficiários em situação de vulnerabilidade de direitos.

No que diz respeito aos métodos quantitativos e qualitativos aplicados, observou-se que os relatórios das análises dos dados devem ser organizados e sistematizados de forma a permitir a recuperação tanto dos dados, quanto dos procedimentos e critérios utilizados para os julgamentos apresentados nos relatórios de avaliação, com vistas a assegurar a consistência e adequação dos métodos, das análises e das conclusões.

Além disso, devem ser instituídos mecanismos de controle para assegurar a integridade dos dados e informações obtidos durante a avaliação, que possibilitem checar erros na coleta, no processamento e nos resultados das análises efetuadas, bem como mecanismos de proteção para garantir a guarda e manutenção das informações coletadas.

7 CONCLUSÕES E RECOMENDAÇÕES

Esta pesquisa teve por objetivo o estudo da utilização das avaliações de programas governamentais no Brasil, enfatizando tanto os aspectos conceituais quanto empíricos. Para tanto, por meio da pesquisa bibliográfica sobre a utilização das avaliações e do levantamento documental sobre as avaliações de programas realizadas no país, buscou-se identificar e examinar os propósitos, a tipologia, os objetivos, as sistemáticas e metodologias empregadas, a quem se destinam e como são utilizados os resultados das avaliações, especialmente aquelas realizadas pela SPI/MPOG, SFC/CGU e TCU. Com base nessas informações, foi verificado em que medida as estratégias e abordagens metodológicas adotadas consideraram as necessidades e prioridades daqueles atores e/ou organizações que devem utilizar os resultados e informações produzidos pelas avaliações para dar cumprimento aos objetivos propostos.

Desta forma, o capítulo 2 apresentou a revisão da bibliografia e a definição do marco conceitual adotado para o desenvolvimento deste estudo, iniciando com a identificação dos propósitos da avaliação, ante a sua importância para o estudo da utilização. Também foram identificados os principais aspectos conceituais a serem aplicados no estudo da utilização das avaliações de programas governamentais, bem como foram estabelecidos os critérios para a análise das avaliações selecionadas para o estudo empírico.

Os propósitos das avaliações de programa condicionam a utilização esperada dos seus resultados, pressupondo a existência de dois elementos: o objetivo da avaliação (para quem?) e os interessados na sua utilização (para quem?).

Assim, para se examinar os propósitos das avaliações, com vistas a analisar se os objetivos foram alcançados – o que pressupõe a utilização das informações – buscou-se, na literatura, os fatores que favorecem a utilização das avaliações, os quais foram agrupados em três categorias principais: aqueles relacionados ao contexto de implementação da avaliação, os relativos ao contexto político no qual as decisões pela realização e execução da avaliação foram tomadas, e a terceira categoria diz respeito ao envolvimento dos *stakeholders* com a avaliação.

Para análise dos fatores que influenciam a utilização, foi necessário identificar um instrumento que possibilitasse examinar, nas avaliações de programas selecionadas para o estudo empírico, a sua observância. Depreendeu-se que os padrões de avaliação de programas definidos pelo *Joint Committee on Standards for Educational Evaluation* refletiam de forma consistente os principais atributos da prática avaliativa e os fatores identificados.

Assim, os padrões de avaliação de programas do *Joint Committee* foram definidos como critérios para examinar as avaliações de programas governamentais no Brasil, com vistas a analisar em que medida as sistemáticas e abordagens metodológicas aplicadas favorecem a utilização dos resultados por elas produzidos, principalmente, no que diz respeito àquelas realizadas pelas organizações públicas responsáveis pela função avaliativa.

Uma vez definidos o arcabouço conceitual e os critérios a serem aplicados no exame da utilização das avaliações no Brasil, procedeu-se ao levantamento, com base na legislação relativa à administração pública federal, dos tipos de avaliação existentes, seus objetivos, características e para quem se destinam, a partir dos quais foram definidos os perfis das avaliações governamentais, conforme apresentado no capítulo 4 e sintetizados na QUADRO 33.

Tipos de Avaliação	Objetivo (para que?)	Interessados Primários na Utilização (Para quem?)	
Avaliação de objetivos estratégicos	Subsidiar a elaboração da LDO	MPOG e Congresso Nacional	
Avaliação de resultados de programas			
Avaliação física e financeira de programas	Aferir o resultado quanto ao cumprimento dos objetivos e das metas fixadas		
	Subsidiar o processo de alocação de recursos		
	Subsidiar a política de gastos públicos		
	Subsidiar a coordenação das ações de governo		
	Evitar a dispersão e o desperdício de recursos públicos		
Avaliação da Qualidade dos Serviços Públicos	Satisfação da sociedade quanto aos bens e serviços ofertados pelo Poder Público		Sociedade
Acompanhamento físico e financeiro dos programas	Subsidiar o processo de alocação de recursos públicos; Subsidiar a política de gastos; Subsidiar a coordenação das ações do governo;		Gestores dos Programas, MPOG, Congresso Nacional
Avaliação da eficácia e efetividade de programas			
Avaliação de investimentos			
Estudos e pesquisas sócio-econômicas	Dar apoio técnico e institucional ao Governo na avaliação, formulação e acompanhamento de políticas públicas, planos e programas de desenvolvimento e oferecer à sociedade elementos para o conhecimento e solução dos problemas e dos desafios do desenvolvimento brasileiro	Gestores dos Programas, MPOG, Congresso Nacional, Sociedade	

(contiuna)

(conclusão)

Tipos de Avaliação	Objetivo (para que?)	Interessados Primários na Utilização (Para quem?)
Avaliação de metas PPA	comprovar a conformidade da sua execução	Congresso Nacional, TCU, Gestores, Sociedade
Avaliação dos orçamentos	comprovar a conformidade da execução com os limites e destinações estabelecidos na legislação pertinente	
Avaliação da execução dos programas	comprovar o nível de execução das metas, o alcance dos objetivos e a adequação do gerenciamento	
Avaliação de eficácia e eficiência da gestão	comprovar a legalidade e a legitimidade dos atos e a examinar os resultados quanto à economicidade, à eficiência e à eficácia da gestão orçamentária, financeira, patrimonial, de pessoal e demais sistemas administrativos e operacionais	
Auditoria Operacional	Obter melhorias em organizações, programas e atividades governamentais	Congresso Nacional, Gestores, Sociedade

QUADRO 33 - TIPOLOGIA DE AVALIAÇÃO DE PROGRAMAS DE GOVERNO NO ÂMBITO DA ADMINISTRAÇÃO PÚBLICA FEDERAL

De posse dessas informações, realizou-se o estudo das sistemáticas e abordagens metodológicas utilizadas para realização das avaliações dos programas do Plano Plurianual 2004-2007, sob a coordenação da Secretaria de Planejamento e Investimentos Estratégicos do Ministério do Planejamento, Orçamento e Gestão; bem como das avaliações de programas do Tribunal de Contas da União, realizadas pela Secretaria de Fiscalização e Avaliação de Programas Governamentais, por intermédio das auditorias operacionais; das avaliações incumbidas à Secretaria Federal de Controle Interno da Controladoria Geral da União, processadas por meio de fiscalizações e auditorias; e ainda, das avaliações decorrentes da Política de Monitoramento e Avaliação do Ministério do Desenvolvimento Social e Combate à Fome, sob a responsabilidade da Secretaria de Avaliação e Gestão da Informação.

Além disso, procedeu-se ao levantamento de informações sobre os relatórios de avaliação publicados; especialmente os relativos ao Plano Plurianual, da SPI/MPOG, das avaliações de programas do Tribunal de Contas e da Controladoria Geral da União e aqueles disponibilizados pelo Ministério do Desenvolvimento Social e Combate à Fome, com vistas a dimensionar o universo a ser estudado e a estratégia metodológica mais adequada para examinar as seguintes questões:

- i) Qual a credibilidade e qualidade das avaliações governamentais de programas?
- ii) Qual a adequação e consistência das abordagens metodológicas adotadas nessas avaliações diante das necessidades daqueles que farão uso das informações?

- iii) Quais os fatores e os mecanismos que influenciam a utilização dos resultados das avaliações?
- iv) Em que medida as abordagens metodológicas adotadas nas avaliações favorecem a sua utilização?

Em vista da diversidade e quantidade de informações e considerando o nível de institucionalização da função avaliativa no Brasil, delimitou-se o período de 2004 a 2007 para o levantamento das informações necessárias à quantificação das avaliações realizadas pelos órgãos acima mencionados.

Assim, com fundamento na revisão de literatura apresentada no capítulo 2 e na análise da relação entre os padrões de avaliação de programas e os fatores que influenciam a utilização e considerando, ainda, as questões propostas na pesquisa empírica, adotou-se, como estratégia metodológica para a análise das avaliações identificadas no capítulo 4, os estudos de casos múltiplos, conforme definido por Yin (2005) e descrito no capítulo relativo à metodologia (3).

Cabe destacar, no entanto, que alguns fatores comprometeram o pleno cumprimento dos objetivos desta pesquisa, principalmente, no que diz respeito à parte empírica, em razão da limitação de prazo para conclusão dos trabalhos, bem como da não disponibilização tempestiva das informações, fatos que impossibilitaram contemplar a análise dos dados das avaliações realizadas pela Secretaria Federal de Controle Interno – SFC/CGU – e pela Secretaria de Avaliação e Gestão da Informação – SAGI/MDS –, muito embora as respectivas sistemáticas de avaliação tenham sido levantadas e apresentadas no capítulo 4, que trata da institucionalização da função avaliação no Brasil.

Em que pese essas limitações, foram coletados e analisados os dados e informações de uma amostra não estatística das avaliações dos programas do PPA 2004-2007 realizadas pela Secretaria de Planejamento e Investimentos Estratégicos – SPI/MPOG – e das avaliações executadas pela Secretaria de Fiscalização e Avaliação de Programas do Tribunal de Contas da União no mesmo período. Com base nessas informações foi possível examinar a adequabilidade das avaliações aos padrões de avaliação definidos pelo Comitê Conjunto, sem a pretensão de se generalizar os achados observados, apenas de evidenciar o nível de aderência à prática avaliativa, com ênfase na utilização, com base nos padrões mencionados, conforme demonstrado nos capítulos 5 e 6.

Como resultado das análises realizadas, pode-se observar que as avaliações de programas governamentais existentes no âmbito da administração pública federal no Brasil

foram instituídas, principalmente, com o objetivo de subsidiar o processo orçamentário e para promover a *accountability*, tanto interna, quanto externa, dos gastos e dos resultados das ações governamentais.

Accountability interna no sentido de diminuir a assimetria de informações entre a burocracia e o parlamento, propiciando um controle mais efetivo destes sobre os primeiros. E, externa mediante a publicidade dos resultados do governo, garantindo-se maior transparência para a sociedade no que diz respeito ao uso dos recursos públicos (MOYNIHAN, 2008).

As três abordagens avaliativas utilizadas neste estudo, as avaliações dos programas do PPA realizadas pela SPI/MPOG, as avaliações de programas do TCU e as fiscalizações e auditorias executadas pela CGU são avaliações externas e evidenciam nos seus objetivos o mesmo propósito, a *accountability*, no entanto elas não compartilham dos mesmos valores. Cada abordagem traz consigo um conjunto diferenciado de critérios, que fazem com que se interceptem, conforme demonstrado na Figura 6, mas não se comuniquem, não há interação entre as avaliações, muito embora elas caminhem paralelamente para a mesma finalidade: aperfeiçoar a administração pública mediante a melhoria da *accountability*, da eficiência da gestão e das decisões orçamentárias.

Uma das explicações para este fato pode ser atribuído à ausência de uma política de avaliação que estabeleça os princípios básicos e os conceitos gerais sobre as dimensões da avaliação. Na ausência de uma política, não há uniformidade de entendimento, sequer em relação ao próprio instrumento de avaliação (TROCHIM, 2009).

Essa constatação pode ser evidenciada pelas diferentes definições adotadas pelas organizações mencionadas para os princípios orientadores da gestão: eficiência, eficácia e efetividade, com base nos quais os resultados deveriam ser examinados.

A eficiência, por exemplo, para o PPA 2004-2007, é a medida da relação entre os recursos efetivamente utilizados para a realização de uma meta para um projeto, atividade ou programa, em relação ao previsto. Para a SFC/CGU, é a medida da relação entre os recursos efetivamente utilizados para a realização de uma meta, frente a padrões estabelecidos. E para as avaliações do TCU, é a relação entre os produtos (bens e serviços) gerados por uma atividade e os custos dos insumos empregados em um determinado período de tempo.

Portanto, a avaliação da eficiência tem diferentes significados segundo o entendimento de cada órgão. Para o PPA 2004-2007, diz respeito à execução financeira; para o SFC/CGU a relação entre os recursos efetivamente utilizados e os padrões estabelecidos; e para o TCU, expressa o custo de uma unidade de produto final em um dado período de tempo.

Desta forma, não é possível compatibilizar os resultados, mesmo porque os resultados das avaliações têm diferentes significados, dificultando, assim, a interpretação e a utilização das informações.

Essa pluralidade conceitual fragiliza a efetiva institucionalização da função avaliação no âmbito da administração pública federal, porque dispersam esforços, enfraquecendo o resultado final desejado, que é contribuir para o aperfeiçoamento das ações governamentais com vistas à melhor prestação dos serviços à sociedade.

Outro fator que dificulta a interação e a comunicação entre as três abordagens avaliativas, a consistência sistêmica, ou seja, aquela desejada para que o sistema de avaliação governamental processe as informações obtidas de várias fontes, e que decorre da ausência de uma política de avaliação, é a falta da definição de padrões que orientem as sistemáticas e as estratégicas metodológicas aplicadas às avaliações no Brasil, de modo a assegurar aos atores envolvidos, e aos interessados nos seus resultados, que as informações e conclusões apresentadas sejam consistentes, assim como os métodos utilizados e as análises são adequados.

Para examinar a consistência interna das avaliações, ou seja, a sistemática e a abordagem metodológica utilizadas nos trabalhos realizados pelos órgãos governamentais responsáveis por essa atividade, com vistas a avaliar a credibilidade e qualidade das avaliações, que são fatores que favorecem a utilização dos estudos, procedeu-se a seleção e análise de uma amostra não estatística das avaliações de programas do PPA e do TCU, realizadas no período de 2004-2007, por intermédio de estudos de casos múltiplos.

Um primeiro desafio para a análise foi a identificação dos padrões a serem utilizados para avaliação. No Brasil, não existem padrões orientadores para a atividade avaliativa. Desta forma, entendeu-se pertinente aplicar os padrões de avaliação de programas estabelecidos pelo Comitê Conjunto de Padrões para Avaliação Educacional dos Estados Unidos, considerando que esses padrões foram instituídos a partir do exercício da atividade avaliativa, seus fundamentos são inerentes à prática da avaliação e são amplamente validados pela literatura pertinente. Desta forma, mesmo não institucionalizados formalmente, os princípios a serem observados, quando da realização desse tipo de trabalho, são tacitamente aceitos.

No que diz respeito às avaliações de programas do PPA 2004-2007, observou-se que os padrões não são aplicáveis, a sistemática e a abordagem metodológica utilizada nessas avaliações não favorecem o julgamento do mérito das informações coletadas. Suas características assemelham-se a dos sistemas de monitoramento. Mas a abordagem metodológica não favorece o acompanhamento efetivo do desempenho, a identificação

tempestiva das deficiências e fragilidades, de forma a permitir a correção oportuna para o alcance dos resultados.

Por outro lado, constatou-se que o processo de avaliação anual dos programas do PPA tem favorecido a disseminação da cultura avaliativa, a promoção do aprendizado organizacional, a melhoria na participação dos gerentes no processo e a percepção da importância do instrumento avaliativo para a melhoria do gerenciamento e da programação orçamentária.

O modelo diálogo interativo de Moynihan (2008) sugere que o aprendizado com base nas avaliações é maior dentro das agências do que em outros ambientes, considerando que existe uma relativa homogeneidade no grupo dos tomadores de decisão que compartilham objetivos similares, tendo constatado que os gestores utilizam as rotinas do diálogo para melhorar a capacidade organizacional e reestruturar os objetivos básicos da organização.

No entanto, o sistema de monitoramento e avaliação adotado pelo Governo Federal carece de aperfeiçoamento, especialmente quanto à definição dos seus objetivos, da sistemática de monitoramento e avaliação e da abordagem metodológica empregada, de forma a assegurar maior qualidade e credibilidade aos resultados apresentados, com vistas a favorecer e ampliar a sua utilização tanto pelos *stakeholders* internos como externos.

Quanto às avaliações do TCU, a análise dos dados evidenciou que os trabalhos realizados pelo Tribunal observam os padrões de avaliação aplicados, com algumas restrições, que podem vir a comprometer a qualidade e credibilidade dos trabalhos.

A abordagem metodológica utilizada pelo TCU para as avaliações examinadas foi estruturada a partir da identificação de problemas que comprometem o desempenho e os resultados esperados para os programas e/ou ações.

Os problemas foram contextualizados por meio do levantamento de informações sobre o programa, que incluem a identificação dos principais *stakeholders* envolvidos ou afetados pelo programa. Com base nas informações coletadas e com a participação direta desses *stakeholders*, foram identificadas as áreas de maior risco e vulnerabilidade para o alcance dos objetivos do programa, que se constituem no escopo dos trabalhos, a partir do qual foram definidas as questões a serem investigadas, como também desenvolvidas, pela própria equipe de avaliação, os instrumentos de coleta de dados.

Uma característica que se observou nas avaliações examinadas do TCU foi o interesse em se obter informações direta e pessoalmente dos beneficiários dos programas, quando possível. Para a boa condução de trabalhos dessa natureza é importante a

normatização dos processos e procedimentos para coleta de dados, principalmente quando investigadas questões que envolvem beneficiários em situação de vulnerabilidade de direitos.

No que diz respeito aos métodos quantitativos e qualitativos aplicados, observou-se que os relatórios das análises dos dados devem ser organizados e sistematizados de forma a permitir a recuperação tanto dos dados, quanto dos procedimentos e critérios utilizados para os julgamentos apresentados nos relatórios de avaliação, com vistas a assegurar a consistência e adequação dos métodos, das análises e das conclusões.

No que diz respeito à utilização das avaliações, a sistemática e a abordagem metodológica adotadas pelo TCU favorecem a utilização dos resultados pelos atores responsáveis pela implementação das recomendações, que são formuladas em todos os relatórios, as quais, inclusive, são monitoradas pelo Tribunal com vistas a assegurar a adoção das medidas corretivas.

O envolvimento dos *stakeholders* no processo de avaliação tem sido apontado na literatura como um dos fatores que mais favorecem a utilização (Johnson *et al.* 2009). Conforme constado por Cousins e Leithwood (1986), o alto envolvimento contribui para elevados níveis de utilização em termos de impacto e na formulação de políticas, assim como favorece o aprendizado organizacional sobre o programa, a programação de decisões a serem feitas e aumenta, em certa medida, outros fatores, como a comunicação, relevância, processamento de informações e a credibilidade.

Da mesma forma que as avaliações de programas do PPA, as avaliações do TCU favorecem a *accountability* das ações governamentais, tanto interna como externa. No entanto, muitas restrições têm sido apresentadas na literatura quanto à efetiva utilidade e utilização dos resultados apresentados por essas avaliações, considerando que nem sempre são suficientes para fornecer as informações necessárias para subsidiar as decisões sobre alocação de recursos e melhorias ou alteração das políticas (WEISS, 2005; RADIN, 2006; PATTON, 2008).

Segundo Radin (2006) as evidências demonstram que as avaliações enfrentam três paradoxos: 1) a ênfase em resultados imensuráveis; 2) as estratégias *on-size-fits-all*; e 3) a expectativa de que a avaliação resolva todos os problemas.

Esses dilemas são observados no caso brasileiro, conforme demonstrou a análise dos dados. As dimensões estabelecidas para a mensuração dos resultados: eficiência, eficácia e efetividade não são sequer compatibilizadas entre as três abordagens examinadas (PPA, TCU e CGU). Ademais, as sistemáticas e abordagens estratégicas utilizadas, em cada caso, podem

ser classificadas como modelo único (*on-size-fits-all*), que descaracteriza as diferenças entre os programas, dando o mesmo tratamento a todos os trabalhos.

Por fim, foi observado que as avaliações governamentais contemplam diversos objetivos para atender diferentes interessados na sua utilização. Segundo Weiss (1998), é impossível atender a todos os propósitos dentro dos limites de um único estudo. Muito embora seja possível contemplar diferentes tipos de questões no mesmo trabalho, nem mesmo um estudo bem planejado poderá fornecer informações para todas as questões apresentadas por todos os atores envolvidos. Há que se considerar, ainda, a possibilidade de incompatibilidade entre os propósitos (WEISS, 1988, p.33).

Patton (2008) ressaltou que diferentes propósitos relacionam-se a diferentes formas de utilização e têm implicações no desenho da avaliação; na mensuração, análise e interpretação dos dados; bem como na forma de apresentação do relatório, disseminação dos resultados e nos critérios para julgamento da qualidade do trabalho.

Portanto, essas considerações nos levam a sugerir algumas áreas para futuras pesquisas, como por exemplo:

- 1) o desenvolvimento de pesquisas com vistas à institucionalização de uma política de avaliação, para o Brasil, que contemple os princípios orientadores da prática avaliativa;
- 2) o desenvolvimento de novas abordagens metodológicas para a avaliação de programas governamentais, considerando os diferentes objetivos a serem alcançados, assim como os atores responsáveis pela sua utilização;
- 3) o desenvolvimento de abordagens metodológicas apropriadas ao estudo da utilização das avaliações; e
- 4) o incentivo à realização de pesquisas voltadas à mensuração da utilização das avaliações de programas governamentais.

Em síntese, espera-se que os resultados desta pesquisa possam contribuir para as discussões sobre a institucionalização da política de avaliação no Brasil, sob a perspectiva da utilização dos seus resultados para a promoção da melhoria social.

REFERÊNCIAS

ALKIN, Marvin C. (Ed). **Evaluation Roots: Tracing Theorists'Views and Influences.** California: Sage Publications, Inc., 2004.

AVALIAÇÃO_ANUAL_PPA2204-2207. Brasília: SPI-MPOG, 2004-2007. Banco de Dados relativo às avaliações dos programas do PPA 2004-2007, realizadas pela SPI por meio da aplicação de questionário.

BARZELAY, Michael. Central Audit Institutions and Performance Auditing: A Comparative Analysis of Organizational Strategies in the OECD. *Governance: An International Journal of Policy and Administration*, Malden, MA, v. 10, n. 3, jul. 1997.

BLAKE, Sarah C.; OTTOSON, Judith M. Knowledge Utilization: implications for evaluations. In: OTTOSON, J. M.; HAWES, P. (Ed.). *Knowledge Utilization, Diffusion, Implementation, Transfer and Translation: implications for evaluations.* New Directions for Evaluation, n. 124, p. 21-34, 2009.

BRANDON, Paul R.; SINGH, J. Malkeet. The Strength of the Methodological Warrants for the Findings of Research on Program Evaluation Use. *American Journal of Evaluation*: v. 30, n. 2, p. 123-157, 2009. Disponível em < <http://online.sagepub.com>>. Acesso em 28/10/2009.

BRASIL. Constituição da República Federativa do Brasil de 1988. Disponível em <http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm>. Brasília, 1988. Acesso em 18/06/2009a.

_____. Lei n ° 8.443, de 16 de julho de 1992. Lei Orgânica do Tribunal de Contas da União. Diário Oficial da União, Brasília, DF, 17 jul. de 1992.

BRASIL. Plano Diretor da Reforma do Aparelho do Estado. Presidência da República, Câmara da Reforma do Estado. Brasília, 1995. Disponível em: http://www.planalto.gov.br/publi_04/COLECAO/PLANDI.HTM. Acesso em: 04/09/2009.

_____. Exposição de Motivos n. 49/95, de 18 de agosto de 1995. Exposição de motivos ao Presidente da República que acompanhou a proposta de Emenda Constitucional que Trata da Reforma Administrativa, encaminhada pelo Governo ao Congresso Nacional, em 23 de agosto de 1997. Brasília, 1997. Disponível em http://www.planejamento.gov.br/gestao/conteudo/publicacoes/reforma_estado/. Acesso em 03/10/2009.

_____. Programa Brasil em Ação. Presidência da República, Brasília, 1998a. Disponível em http://www.planalto.gov.br/publi_04/COLECAO/2ACAO.HTM. Acesso em 07/09/2009.

_____. Decreto n. 2.829, de 29 de outubro de 1998. Estabelece normas para a elaboração e execução do Plano Plurianual e dos Orçamentos da União e dá outras providências, Brasília, 1998b. Disponível em: < http://www.planalto.gov.br/ccivil_03/decreto/D2829.htm>. Acesso em 04/09/2009.

_____. Tribunal de Contas da União. Relatório Grupo de Pesquisa Tendências do Controle da Gestão Pública: paradigmas do controle externo. Brasília, 1999.

_____. Tribunal de Contas da União. Portaria Normativa 144 de 10 de julho de 2000. Manual de Auditoria de Natureza Operacional. Boletim do TCU nº 37/2000, Brasília, DF, 2000.

_____. Lei n. 10.180, de 6 de fevereiro de 2001. Organiza e disciplina os Sistemas de Planejamento e de Orçamento Federal, de Administração Financeira Federal, de Contabilidade Federal e de Controle Interno do Poder Executivo Federal, e dá outras providências. Brasília, 2001a. Disponível em: http://www.planalto.gov.br/ccivil_03/Leis/LEIS_2001/L10180.htm. Acesso em 07/09/2009.

_____. Secretaria Federal de Controle Interno. Instrução Normativa n. 01 de 6 de abril de 2001. Brasília, 2001b. Disponível em < http://www.cgu.gov.br/Legislacao/Arquivos/InstrucoesNormativas/IN01_06abr2001.pdf>. Acesso em 09/02/2010.

_____. Tribunal de Contas da União. Resolução n. 155, de 4 de dezembro de 2002. Aprova o Regimento Interno do Tribunal de Contas da União. Diário Oficial da União, Brasília, DF, 09 dez. 2002. Seção 1, p. 125. Disponível em < http://portal2.tcu.gov.br/portal/page/portal/TCU/normativos/BTCU_ESPECIAL_01_DE_13_02_2007.pdf>. Acesso em: 01/10/2009.

_____. Decreto nº 4.745, de 16 de junho de 2003. Aprova o Estatuto e o Quadro Demonstrativo dos Cargos em Comissão e das Funções Gratificadas do Instituto de Pesquisa Econômica Aplicada - IPEA, e dá outras providências. Diário Oficial da União, Brasília, DF, 17 de jun. 2003. Disponível em < http://www.planalto.gov.br/ccivil_03/decreto/2003/D4745.htm>. Acesso em 09/02/2010.

_____. Lei nº 10.933, de 11 de agosto de 2004. Plano Plurianual para o período 2004/2007. Diário Oficial da União, Brasília, DF, 12 ago. de 2004a.

_____. Decreto n. 5.233, de 6 de outubro de 2004. Estabelece normas para a gestão do Plano Plurianual 2004-2007 e de seus Programas e dá outras providências. Diário Oficial da União, Brasília, DF, 07 out. 2004b.

_____. Ministério do Planejamento Orçamento e Gestão. Manual de Avaliação Anual do Plano Plurianual 2004-2007: Ano Base 2004. Brasília, 2004c.

_____. Tribunal de Contas da União. Relatório de Avaliação: Assistência Financeira para Aquisição e Distribuição de Medicamentos Excepcionais. Brasília, 2005a.

_____. Tribunal de Contas da União. Relatório de Avaliação: Programa Combate ao Abuso e Exploração Sexual de Crianças e Adolescentes. Brasília, 2005b.

_____. Relatório Anual de Avaliação do Plano Plurianual 2004-2007: exercício 2005 ano base 2004. Ministério do Planejamento, Orçamento e Gestão, Secretaria de Planejamento e Investimentos Estratégicos, Brasília, 2005c. Disponível em <http://www.sigplan.gov.br/v4/appHome/>. Acesso em: 01/09/2009.

_____. Ministério do Planejamento Orçamento e Gestão. Manual de Avaliação do Plano Plurianual 2004-2007 Exercício 2006, Ano Base 2005. Brasília, 2006a.

_____. Relatório Anual de Avaliação do Plano Plurianual 2004-2007: exercício 2006 ano base 2005. Ministério do Planejamento, Orçamento e Gestão, Secretaria de Planejamento e Investimentos Estratégicos, Brasília, 2006b. Disponível em <http://www.sigplan.gov.br/v4/appHome/>. Acesso em: 01/09/2009.

_____. Tribunal de Contas da União. Relatório de Avaliação: Programa Segurança Fitozoossanitária de Produtos Agropecuários. Brasília, 2006c.

_____. Tribunal de Contas da União. Relatório de Avaliação: Programa Segundo Tempo. Brasília, 2006d.

_____. Tribunal de Contas da União. Relatório de Avaliação: Programa Nacional de Erradicação da Hanseníase. Brasília, 2006e.

_____. Tribunal de Contas da União. Relatório de Avaliação: Programa Doação Captação e Transplante de Órgãos e Tecidos. Brasília, 2006f.

_____. Tribunal de Contas da União. Relatório de Avaliação: Programa Manutenção de Hidrovias. Brasília, 2006g.

_____. Tribunal de Contas da União. Relatório de Avaliação: Programa Governo Eletrônico. Brasília, 2006h.

_____. Ministério do Desenvolvimento e Combate à Fome. Portaria nº 329 de 11 de outubro de 2006. Política de Monitoramento e Avaliação do Ministério do Desenvolvimento e Combate à Fome. Brasília, 2006i. Disponível em <<http://www.mds.gov.br/institucional/secretarias/secretaria-de-avaliacao-e-gestao-da-informacao-sagi/arquivo-sagi/portarias>>. Acesso em 18/09/2009.

_____. Tribunal de Contas da União. Relatório de Auditoria de Natureza Operacional nas Ações de Fiscalização e de Acompanhamento de Ingressos em Terras Indígenas. TC nº 012.484/2005-6. Brasília, 2007a.

_____. Tribunal de Contas da União. Relatório de Auditoria de Natureza Operacional em Ações do Programa Desenvolvimento Sustentável da Aquicultura. TC nº 017.237/2006-6. Brasília, 2007b.

_____. Relatório Anual de Avaliação do Plano Plurianual 2004-2007: exercício 2007 ano base 2006. Ministério do Planejamento, Orçamento e Gestão, Secretaria de Planejamento e Investimentos Estratégicos, Brasília, 2007c. Disponível em <http://www.sigplan.gov.br/v4/appHome/>. Acesso em: 01/09/2009.

_____. Ministério do Planejamento Orçamento e Gestão. Orientações Gerais para registro de informações no Módulo de Avaliação do Sigplan: Exercício 2008, Ano-Base 2007. Brasília, 2008a. Disponível em <http://www.sigplan.gov.br/v4/appHome/>. Acesso em: 01/09/2009.

_____. Relatório Anual de Avaliação do Plano Plurianual 2004-2007: exercício 2008 ano base 2007. Ministério do Planejamento, Orçamento e Gestão, Secretaria de Planejamento e Investimentos Estratégicos, Brasília, 2008b. Disponível em <http://www.sigplan.gov.br/v4/appHome/>. Acesso em: 01/09/2009.

_____. Tribunal de Contas da União. Relatório de Auditoria de Natureza Operacional a Ação Levantamentos Hidrogeológicos do Programa Geologia do Brasil – 1115. TC nº 018.202/2006. Brasília, 2008c.

_____. Tribunal de Contas da União. Relatório de Auditoria de Natureza Operacional Fundação Nacional de Saúde – FUNASA (UG255000). TC nº 009.240/2007-5. Brasília, 2008d.

CALMON, Kátya M. N.; GUSSO, Divonzir A. A Experiência de Avaliação do Plano Plurianual (PPA) do Governo Federal no Brasil. In: OFICINA DE TRABALHO DO PROJETO FORTALECIMENTO DA FUNÇÃO AVALIAÇÃO NA AMÉRICA LATINA, PROMOVIDA POR IPEA E CEPAL. 2003, Santiago.

CHELIMSKY, Eleanor. The Politics of Program Evaluation. *New Directions for Program Evaluation*, San Francisco, n. 34, p. 5-21, 1987. Disponível em < <http://www3.interscience.wiley.com/journal/117946202/grouphome/home.html>>. Acesso em 11/11/2009.

CHELIMSKY, Eleanor. Integrating Evaluation Units into the Political Environment of Government: the role of evaluation policy. In TROCHIM, W. M. K.; MARK, M. M.; COOKSY, L. J. (Ed.). *Evaluation Policy and Evaluation Practice*. *New Directions for Evaluation*, n. 123. P. 51-66, 2009. Disponível em < www.aje.sagepub.com >. Acesso em 20/02/2009.

COUSINS, J. Bradley. Commentary: Minimizing Evaluation Misuse as Principled Practice. *American Journal of Evaluation*, v. 25, n. 3, p. 391-397, 2004. Disponível em < www.aje.sagepub.com >. Acesso em 20/02/2009.

COUSINS, J. Bradley; LEITHWOOD, Kenneth A. Current Empirical Research on Evaluation Utilization. *Review of Educational Research*, v. 56, n. 3, p. 331-364, 1986. Disponível em <rer.aera.net>. Acesso em 26/06/2009.

DERLIEN, Hans-Ulrich. Genesis and Structure of Evaluation Efforts in Comparative Perspective. In: RIST, Ray C. (Ed.). **Program Evaluation and the Management of Government: patterns e prospects across eight nations**. New Brunswick: Transaction Publishers, 1990.

DFID. *Equidade, Controle Social e Governança na Gestão Pública*. Brasil: DFID, 2006.

EUA. General Accounting Office. *Standards for Audit of Governmental Organizations, Programs, Activities and Functions*. Tradução de: SILVA, J. D. Wolff da. Brasília, 1984.

FLEISCHER, Dreolin N.; CHRISTIE, Christina A. Evaluation Use: results from a survey of U.S. American Evaluation Association. *American Journal of Evaluation*, v. 30, n. 2, p. 158-175, 2009. Disponível em < aje.sagepub.com >. Acesso em 28/10/2009.

GARCIA, Ronaldo C. *Subsídios para Organizar Avaliações da Ação Governamental*. Texto para Discussão, Brasília: IPEA, n. 776, 2001.

HENRY, Gary T.; MARK, Melvin M. Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation*, v. 24, n. 3, p. 293-314, 2003. Disponível em < aje.sagepub.com >. Acesso em 20/02/2009.

_____. Influential Evaluations. *American Journal of Evaluation*, v. 24, n. 4, p. 515-524, 2003. Disponível em < www.aje.sagepub.com >. Acesso em 20/02/2009.

KETTL, Donald F. *The Transformation of Governance: Public Administration for Twenty-First Century America*. Baltimore, Maryland: The John Hopkins University Press, 2002.

INTOSAI. *Implementation Guidelines For Performance Auditing*. Estocolmo, jul. 2004.

JOHNSON, R. Burke. Toward a Theoretical Model of Evaluation Utilization. *Evaluation and Program Planning*, v. 21, p. 93-110, 1998.

JOHNSON, Kelli *et al.* Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, v. 30, n. 3, p. 377-410, 2009. Disponível em < aje.sagepub.com >. Acesso em 28/10/2009.

JOINT COMMITTEE FOR EDUCATIONAL EVALUATION. Disponível em <http://www.jcsee.org/about>. Acesso em: 22/01/2010.

KIRKART, Karen. Reconceptualizing Evaluation Use: An Integrated Theory of Influence. *New Directions for Evaluation*. CARACELLI, V.J.; PRESKILL, H. (Ed). San Francisco, n. 88, p. 5-23, 2000.

KUSEK, Jody Z; RIST, Ray C. **Ten Steps to a Results-Based Monitoring and Evaluation System**. Washington: The World Bank, 2004.

LAWRENZ, Frances; GULLICKSON, Arlen; TOAL, Stacie. Dissemination: Handmaiden to Evaluation Use. *American Journal of Evaluation*, v. 28, n. 3, p. 275-289, 2007. Disponível em < www.aje.sagepub.com >. Acesso em 20/02/2009.

LAWRENZ, Frances *et al.* Relationships between Involvement and Use in the Context of Multi-site Evaluation. Conferência da Associação Americana de Avaliação, 12/11/2009, Orlando, Flórida. Disponível em < <http://www.eval.org/eval2009/default.htm> >. Acesso em 15/12/2009.

LEVITON, Laura; Hughes, Edward F. X. Research On the Utilization of Evaluations: A Review and Synthesis. *Evaluation Review*, v. 5, n. 4, p. 525-548, 1981. Disponível em < www.aje.sagepub.com >. Acesso em 20/02/2009.

_____. Evaluation Use: Advances, Challenges and Applications. *American Journal of Evaluation*, v. 24, n. 4, p. 525-535, 2003. Disponível em < www.aje.sagepub.com >. Acesso em 20/02/2009.

LYNN, Laurence E. *Public Management: old and new*. New York, NY: Routledge, 2006.

LOVE, Arnold; RUSSON Craig. Evaluation Standards in an Internacional Context. *New Directions for Evaluation*, n. 104, p. 5-14, 2004. Disponível em < <http://www3.interscience.wiley.com/journal/117946202/group/home/home.html>>. Acesso em 11/11/2009.

MARCH, James G. **A Primer on Decision Making: how decisions happen**. Chip Heath (Assist.). New York: The Free Press, 1994.

MARK, Melvin M.; HENRY, Gary T. The Mechanisms and Outcomes of Evaluation Influence. *Evaluation*, v. 10, p. 35-56, 2004. Disponível em < www.evi.sagepub.com >. Acesso em 26/03/2009.

MARK, Melvin M.; HENRY, Gary T.; JULNES, George. **Evaluation: an integrated framework for understanding, guiding, and improving policies and programs**. San Francisco: Jossey-Bass, 2000.

MARRA, Mita. How Much Does Evaluation Matter?: Some Examples of the Utilization of the Evaluation of the World Bank's Anti-Corruption Activities. *Evaluation*, v. 6, p. 22-36. Disponível em < www.evi.sagepub.com >. Acesso em 23/01/2009.

MEER, Frans-Bauke V.; EDELENBOS, Jurian. Evaluation in Multi-Actor Policy Process: Accountability, Learning and Co-operation. *Evaluation*, v. 12, p. 201-218, 2006. Disponível em < www.evi.sagepub.com >. Acesso em 26/03/2009.

MOYNIHAN, Donald P. *The Dynamics of Performance Management: constructing information and reform*. Washington, District of Columbia: Georgetown University Press, 2008.

PADILHA, Cláudia de O. **Uma Leitura Crítica das Prioridades Definidas no PPA 2004-2007 a partir do Estudo Comparativo das Metas Fixadas nas LDOs e LOAs**. 151 f. Monografia (Especialização em Orçamento Público) - Instituto Serzedello Côrrea/TCU, Centro de Formação, Treinamento e Aperfeiçoamento/Câmara dos Deputados e Universidade do Legislativo Brasileiro/Senado Federal, Brasília, 2008.

PATTON, Michael Q. **Utilization-Focused Evaluation**. 4. ed. California: Sage Publications, Inc., 2008.

_____. Reports on Topic Areas: The Evaluator's Responsibility for Utilization. *American Journal of Evaluation*, v. 9, p. 5-24, 1988. Disponível em < www.aje.sagepub.com >. Acesso em 9/02/2009.

PRESKILL, Hallie; CARACELLI, Valerie. Current and Developing Conceptions of Use: Evaluation Use TIG Survey Results. *Evaluation Practice*, v. 18, n. 3, p. 209-225, 1997. Disponível em < www.aje.sagepub.com >. Acesso em 20/02/2009.

_____; ZUCKERMAN, Barbra; MATTHEWS, Bonya. An Exploratory Study of Process Use: Findings and Implications for Future Research. *American Journal of Evaluation*, v. 24, n. 4, p. 423-442, 2003. Disponível em < www.aje.sagepub.com >. Acesso em 26/03/2009.

POISTER, Theodore H. Performance Monitoring. In: WHOLEY, Joseph S. *et al.* **Handbook of Practical Program Evaluation**. 2 Ed. San Francisco: Jossey-Bass, 2004. p. 98-125.

RADIN, Beryl A. **Challenging the Performance Movement: Accountability, Complexity and Democratic Values**. Washington, District of Columbia: Georgetown University Press, 2006.

RIST, Ray C. (Ed.). **Program Evaluation and the Management of Government: patterns e prospects across eight nations**. New Brunswick: Transaction Publishers, 1990.

SANDERS, James R. **The Program Evaluation Standards: how to assess evaluations of educational programs**. 2. ed. California: Sage Publications, 1994.

SERPA, Selma M. H. C. **Sistemática de Classificação de Programas para Seleção de Objetos de Fiscalização com vistas à Avaliação do Desempenho da Ação Governamental**. 107f. Monografia Curso de Especialização em Controle Externo, Instituto Serzedello Correa, Tribunal de Contas da União, Brasília, 2004. Disponível em < <http://www.ticontrole.gov.br/portal/pls/portal/docs/690430.PDF> >. Acesso em 09/02/2010.

_____; MARTINS, Eliane V. Olhar de Fora: instrumento de fiscalização do TCU avalia o desempenho da ação estatal. *SP.GOV*, São Paulo, Ano 2, n. 6, p. 38-41, 2005.

SILVA, Pedro L. B.; COSTA, Nilson R. A avaliação de programas públicos: reflexões sobre a experiência brasileira. Brasília: IPEA, 2002. 63 p. Relatório técnico.

SHULHA, Lyn M.; COUSINS, J. Bradley. Evaluation Use: Theory, Research, and Practice Since 1986. *Evaluation Practice*, v. 18, n. 3, p. 195-208, 1997. Disponível em < aje.sagepub.com >. Acesso em 20/02/2009.

STUFFLEBEAM, Daniel L. Empowerment Evaluation, Objectivist Evaluation, and Evaluation Standards: Where the Future of Evaluation Should Not Go and Where It Needs to Go. *Evaluation Practice*, v. 15, n. 3, p. 321-338, 1994. Disponível em < WWW.aje.sagepub.com >. Acesso em: 31/03/2009.

TORRES, Rosalie T.; PRESKILL, Hallie. Evaluation and Organizational Learning: Past, Present and Future. *American Journal of Evaluation*, v. 22, n. 3, p. 387-395, 2001. Disponível em < www.aje.sagepub.com >. Acesso em 26/03/2009.

TROCHIM, William M. K. Evaluation Policy and Evaluation Practice. *New Directions for Evaluation*, n. 123, p. 13-32, 2009. Disponível em < <http://www3.interscience.wiley.com/journal/117946202/grouphome/home.html> >. Acesso em 28/10/2009.

VAITSMAN, Jeni; RODRIGUES, Roberto W. S.; SOUSA, Rômulo. **O Sistema de Avaliação das Políticas Sociais: a experiência do Ministério do Desenvolvimento e Combate à Fome do Brasil**. Brasil: UNESCO, 2006.

VAN EVERA, Stephen. **Guide to Methods for Students of Political Science**. Ithaca: Cornell University Press, 1997.

WEISS, Carol H. Reports on Topic Areas: Evaluation for Decisions: Is Anybody There? Does Anybody Care? *American Journal of Evaluation*, v. 9, p. 5-19, 1988a. Disponível em < www.aje.sagepub.com >. Acesso em 9/02/2009.

_____. If Program Decisions Hinged Only on Information: A Response to Patton. *American Journal of Evaluation*, v. 9, p. 15-28, 1988b. Disponível em < www.aje.sagepub.com >. Acesso em 9/02/2009.

_____. Where Politics and Evaluation Meet. *Evaluation Practice*, v. 14, n. 1, p. 93-106, 1993. Disponível em < www.aje.sagepub.com >. Acesso em 31/03/2009.

_____. **Evaluation**. 2. Ed. Upper Saddle River: Prentice-Hall, Inc., 1998.

_____. Have We Learned Anything New About the Use of Evaluation? American Journal of Evaluation, v. 19, n. 1, p. 21-33, 1998a. Disponível em < www.aje.sagepub.com >. Acesso em 9/02/2009.

_____; MURPHY-GRAHAM, Erin; BIRKELAND, Sarah. An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E. American Journal of Evaluation, v. 26, n. 12, p. 12-30, 2005. Disponível em < www.aje.sagepub.com >. Acesso em 20/02/2009.

_____ *et al.* The Fairy Godmother and Her Warts: Making the Dream of Evidence-Based Policy Come True. American Journal of Evaluation, v. 29, p. 29-47, 2008. Disponível em < www.aje.sagepub.com >. Acesso em 20/02/2009.

WHORTHEN, Blaine R.; SANDERS, James R.; FITZPATRICK, Jody L. **Avaliação de Programas: concepções e práticas**. AZEVEDO, D. A. (Tr.). São Paulo: Editora Gente, 2004.

YIN, Robert. K. **Estudo de Caso: planejamento e métodos**. 3. ed. GRASSI, Daniel (TR). São Paulo: Bookman, 2005.

APÊNDICE

QUESTIONÁRIO – AVALIAÇÃO DA OBSERVÂNCIA DOS PADRÕES DE AVALIAÇÃO DEFINIDOS PELO *JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION*

Nome:

Auditoria:

Para melhor compreensão dos padrões adicionamos na coluna à esquerda algumas orientações explicativas.

Considerando a escala abaixo, indique, com base na auditoria relacionada, realizada sob sua coordenação, se foram observados os padrões objeto das questões a seguir atribuindo a nota que melhor represente a sua percepção, levando em consideração toda a documentação disponível sobre o trabalho realizado.

Caso a questão não seja compreendida, por favor, marque "?" e acrescente a sua dúvida e/ou comentário

Escala	Especificação
NA	Não aplicável
1	Não observado
2	Observado com restrições que limitam o pleno atendimento da questão examinada
3	Observado com restrições, mas atende à questão examinada
4	Observado

Padrões

Utilidade

Orientações	Questões	Nota	Comentário
Pessoas envolvidas na avaliação ou por ela afetadas devem ser identificadas para que suas necessidades possam ser contempladas.	U1 - Identificação Stakeholder		
	Os destinatários da avaliação foram identificados?		
	Os stakeholders relevantes para definição do problema foram identificados?		
	Existe um processo (rotina) para identificação dos stakeholders?		
	Existe a definição das atribuições de cada stakeholder?		
	Os stakeholders identificados participaram do processo de desenho da avaliação?		
	Os stakeholders participaram do processo de execução da avaliação?		
	As demandas de informações dos principais stakeholders foram atendidas?		
Pessoas conduzindo avaliação devem ser competentes e confiáveis para a sua realização.	U2 - Credibilidade do Avaliador		
	Existem critérios para selecionar os avaliadores?		
	A equipe de avaliadores e o plano de trabalho foram avaliados?		
	O Plano da avaliação foi apresentado para os diversos stakeholders e foi considerando realístico e tecnicamente adequado para responder às demandas apresentadas pelos stakeholders?		
	Foram identificados os atores-chave para acompanhar o progresso dos trabalhos de avaliação?		

(continua)

(continuação)

Orientações	Questões	Escala	Comen- tário
As avaliações precisam ser relevantes para os objetivos dos tomadores de decisão, importantes para os stakeholders significativos e suficientemente compreensível para suportar o julgamento quanto ao mérito ou sucesso do programa.	<p>U3 - Informações Necessárias para Definição e Seleção do Escopo</p> <p>As necessidades do requerente da avaliação (cliente) foram compreendidas pelo avaliador? (v. U1)</p> <p>Foram consultados representantes dos principais stakeholders para compreensão das suas perspectivas e identificação de possíveis conflitos de pontos de vista e de necessidade de informações?</p> <p>O requerente da avaliação classificou por ordem de importância os públicos destinatários e buscou identificar a ordem de importância das questões apresentadas por esses grupos a serem tratadas nas avaliações?</p>		
As perspectivas, procedimentos e racionalidade utilizada para interpretar os achados devem ser cuidadosamente descritos, tendo em vista que as bases para o julgamento de valor devem ser claras.	<p>U4 - Identificação dos Valores (Critérios)</p> <p>Existem critérios definidos para o julgamento dos achados da avaliação?</p>		
O relatório da avaliação deveria descrever com clareza o programa avaliado, incluindo o seu contexto, propósitos, procedimentos e achados da avaliação, de forma que as informações essenciais sejam apresentadas e facilmente compreendidas.	<p>U5 - Clareza do Relatório</p> <p>Existem formas de apresentação do relatório considerado as peculiaridades dos públicos identificados?</p> <p>Os resultados das avaliações foram apresentados de forma adequada para os diferentes públicos?</p> <p>A contextualização do programa, bem como as informações disponibilizadas pela avaliação são consistentes com as conclusões e recomendações?</p> <p>O relatório foi submetido à consideração do requerente da avaliação, assim como dos representantes dos principais stakeholders (indicados previamente) antes da sua disponibilização final?</p>		
Os achados intermediários significativos e o relatório da avaliação devem ser disseminados para os interessados na sua utilização, possibilitando a sua tempestiva utilização (os interessados na utilização são aqueles que devem ser informados sobre os resultados da avaliação).	<p>U6 - Tempestividade e Disseminação do Relatório</p> <p>Os proponentes da avaliação e os representantes dos grupos chaves de stakeholders deixaram claro seus principais objetivos com relação à avaliação, a forma de apresentação do relatório, e a data mais apropriada para sua apresentação? (v. U1; U3; U5)</p> <p>Houve alguma negociação quanto ao plano de disseminação da avaliação para os interessados na sua utilização e o acesso deles aos achados e recomendações específicas?</p> <p>A avaliação foi submetida a algum controle de qualidade e seu resultado foi apresentado aos interessados na utilização?</p>		

(continua)

(conclusão)

Orientações	Questões	Nota	Comen- tário
As avaliações devem ser planejadas, conduzidas e relatadas de forma que encoraje o seu acompanhamento posterior pelos stakeholders, assim a probabilidade da avaliação ser utilizada será maior. O impacto da avaliação diz respeito à influência que teve para as decisões e acompanhamento das ações pelos membros dos públicos destinatário.	U7 - Impacto Avaliação		
	Houve alguma demonstração no início da avaliação de como os seus achados poderiam ser úteis para o trabalho desenvolvido?		
	De alguma forma houve o envolvimento dos stakeholders na definição das questões a ser avaliadas e no planejamento e execução da avaliação?		
	Os resultados intermediários foram apresentados de forma periódica?		

Viabilidade

Orientações	Questões	Escala	Comen- tário
Os procedimentos da avaliação devem ser práticos, mantendo o mínimo de interrupções enquanto as informações necessárias são obtidas. Os procedimentos da avaliação são as ações adotadas no processo de coleta e utilização das informações para julgar o sucesso ou mérito do programa.	F1 - Procedimentos Práticos		
	A equipe responsável pela coleta dos dados foi treinada e qualificada para realizar os trabalhos?		
	Os procedimentos de coleta de dados foram compatíveis com o nível de capacitação oferecido?		
	Os procedimentos foram selecionados considerando as limitações de tempo e a disponibilidade de participação dos respondentes?		
A avaliação deve ser planejada e conduzida antecipando-se as diferentes posições dos vários grupos interessados, de forma a se obter a cooperação de todos e assim possivelmente evitar que qualquer desses grupos restrinja as atividades da avaliação, como também prevenindo a utilização inadequada dos seus resultados.	F2 - Viabilidade Política		
	Os avaliadores tiveram acesso aos dados necessários e controle sobre a edição e disseminação de seus relatórios?		
	F3 - Custo-Efetividade		
	Houve alguma avaliação do custo e dos benefícios da avaliação?		

Propriedade

Orientações	Questões	Escala	Comen- tário
As avaliações devem ser concebidas para auxiliar as organizações a conduzir e efetivamente atender às necessidades do público alvo.	P1 - Orientação para Prestação dos Serviços		
	Os stakeholders foram informados dos propósitos da avaliação? (v. A3)		
	O foco da avaliação foi nas características do programa que mais afetam os beneficiários e que promovem o alcance dos objetivos da organização? (v. U3;F3)		
	Foram apresentados resultados intermediários citando forças e deficiências e sugestões para aperfeiçoamento?		

(continua)

(continuação)

Orientações	Questões	Escala	Comen- tário
	<p>P2 - Acordo Formal</p> <p>A avaliação foi contratada?</p> <p>No termo de formalização da contratação, os propósitos da avaliação foram definidos?</p> <p>No termo de contrato, os produtos a serem entregues foram especificados?</p> <p>No termo de contrato estão especificados, os procedimentos de coleta de dados, incluindo: fontes de informações, seleção de participantes; instrumentos e outras técnicas de obtenção de dados, e ainda, se apropriado, o plano de visitas in loco.</p> <p>No termo de contrato, os procedimentos para análise quantitativa e qualitativa de dados e informações e um plano de gerenciamento foram definidos?</p> <p>No termo de contrato, o plano de entrega de relatórios, incluindo considerações sobre os relatórios intermediários e final e formatos de apresentação; anonimidade dos respondentes, prerelease, revisão do relatório, contraprova daqueles sendo avaliados; e editorial autorização de release final sobre os relatórios completos foram definidos?</p> <p>No termo de contrato, os métodos de controle de qualidade da coleta, análise e relatório foram estabelecidos?</p> <p>No termo de contrato, foram definidos os serviços, pessoas, informações e materiais a serem disponibilizados pelo cliente (proponente/requerente) incluindo acesso aos dados?</p> <p>No termo de contrato, foi definido cronograma dos trabalhos tanto clientes como avaliadores?</p> <p>O termo de contrato especifica o orçamento e formas de pagamento?</p> <p>Há previsão de período para revisão e/ou modificação definido no termo de contrato?</p>		
<p>As avaliações devem ser concebidas e conduzidas respeitando e protegendo os assuntos relativos aos direitos e bem-estar humanos.</p>	<p>P3 - Questões Relativas a Direitos Individuais</p> <p>Os valores e a cultura dos participantes foram devidamente compreendidos pelos avaliadores?</p> <p>Existe algum documento assegurando que os direitos dos participantes serão protegidos?</p> <p>Existe permissão formal das autoridades para o acesso aos dados individuais (registros)?</p> <p>Existe autorização dos pais e responsáveis para que crianças sejam utilizadas como fonte de informação/avaliadas?</p> <p>Existem procedimentos que garantam a confidencialidade dos participantes que fornecem informações para serem utilizadas na avaliação?</p> <p>A identidade dos participantes foi revelada nos achados da avaliação?</p> <p>Existem garantias para que outras partes não utilizem os dados coletados para outros fins que não os acordados entre as pessoas que forneceram os dados?</p>		

(conclusão)

Orientações	Questões	Escala	Comentário
Os avaliadores devem respeitar a dignidade e valores em suas relações com outras pessoas associadas com a avaliação, de forma que os participantes não sejam ameaçados ou ofendidos.	<p>P4 - Relações Humanas</p> <p>A cultura, os valores sociais e as diferenças de linguagem dos participantes foram devidamente compreendidos? v. U4, P1 e P3</p> <p>Existem canais de comunicação estabelecidos com os participantes da avaliação?</p> <p>Os avaliadores demonstram familiaridade com a organização onde os trabalhos estão sendo realizados?</p>		
A avaliação deve ser completa e imparcial no exame e registro das forças e fraquezas do programa avaliado, assim pode-se avançar nos pontos fortes e encaminhar as áreas com problemas.	<p>P5 - Avaliação (julgamento) completa e imparcial (justa)</p> <p>O relatório aponta tanto as fraquezas quanto as forças (intencionais e não intencionais) com a devida justificativa? V. U5 e A11</p> <p>O Relatório foi submetido à revisão dos pares quanto a sua consistência e imparcialidade?</p> <p>Algum dado deixou de ser levantado em função do tempo e do custo para obtê-lo? Esta omissão consta do Relatório? Seus efeitos foram considerados?</p>		
As partes formais de uma avaliação devem assegurar que a totalidade dos achados com as suas pertinentes limitações estarão acessíveis para as pessoas afetadas pela avaliação e quaisquer outras com expressos direitos legais de receber os resultados.	<p>P6 - Apresentação dos achados</p> <p>Existe um termo formal definindo os direitos e deveres do contratante (requerente) e do avaliador assegurando o direito de conhecer todos os requerimentos, incluindo identificação dos stakeholders intermediários e finais, autoridade para editar os relatórios, documentação dos propósitos, procedimentos e resultados, quando, onde e que informação sobre a avaliação será divulgada? v. P2</p> <p>O relatório da avaliação foi apresentado de forma completa, com todos os achados e sem omissões ou alterações?</p> <p>A relação entre os propósitos da avaliação, os métodos utilizados, a coleta de dados e os achados são claramente percebidos? V. A10</p> <p>Os julgamentos e recomendações representam perspectivas amplas, balanceadas e evidenciadas no relatório?</p>		
Os conflitos de interesse devem ser tratados com clareza e honestidade de forma a não comprometer o processo e os resultados da avaliação.	<p>P7 - Conflitos de Interesses</p> <p>É possível identificar vantagens (financeiras, social, moral ou política) que as partes poderiam ganhar ou perder com os resultados da avaliação?</p> <p>Existe algum acordo que proteja contra os problemas associados aos conflitos de interesse?</p>		
A alocação de recursos e as despesas realizadas pelo avaliador devem estar refletidas nos procedimentos de prestação de contas, caso contrário, deve-se ser prudente e eticamente responsável, de forma que as despesas sejam comprovadas e apropriadas.	<p>P8 - Responsabilidade Fiscal</p>		

Acuracidade

Orientações	Questões	Escala	Comen- tário
<p>O programa avaliado deve ser descrito e claramente documentado com acuracidade, de forma a que o programa seja claramente identificado.</p>	<p>A1 - Documentação do Programa A documentação relativa à avaliação está disponível? Onde? É possível identificar nos registros a preocupação do avaliador em conhecer e documentar o levantamento das informações sobre o programa, tanto os seus propósitos, como a sua forma de implementação, quanto às características do: pessoal, custos, procedimentos, localização, facilidades, organização, atividades, objetivos, natureza da participação stakeholders e possíveis efeitos colaterais.</p>		
<p>O contexto em que o programa está inserido deve ser examinado com bastante detalhe, com vistas a identificar a probabilidade de sua influência sobre o programa avaliado.</p> <p>O contexto inclui a localização geográfica do programa, duração (existência), clima social e político em torno do programa, atividades concorrentes em andamento, pessoal, condições econômicas.</p>	<p>A2 - Análise do Contexto Os contextos técnico, social, político, organizacional e econômico foram descritos? Foram citadas as fontes para descrição dos contextos? Existe registro de circunstâncias não-usuais que podem ter influenciado os achados?</p>		
<p>Os propósitos e procedimentos da avaliação devem ser monitorados e descritos em detalhes, de forma que possam ser identificados e avaliados</p> <p>Os propósitos da avaliação são definidos em termos dos objetivos da avaliação e na intenção de utilização de seus resultados (para quê? Para quem?)</p> <p>Os procedimentos incluem a forma com que os dados e informações são obtidos, organizados, analisados e reportados para atender aos propósitos da avaliação</p>	<p>A3 - Descrição dos Propósitos e Procedimentos Os propósitos e as intenções de utilização dos achados foram inicialmente definidos? Houve alterações na proposta inicial? As alterações foram registradas? Os papéis de trabalho (notas de campo) e procedimentos para coleta e análise de dados e informações estão disponíveis para revisão e reanálise? Houve monitoramento desse processo por um avaliador independente?</p>		
<p>As fontes de informações utilizadas na avaliação devem ser descritas em detalhes, de forma a possibilitar a avaliação da informação.</p> <p>Os avaliadores deveriam documentar, justificar e relatar suas fontes de informações, os critérios e métodos utilizados para selecioná-los, os meios utilizados para obter as informações dessas fontes e qualquer outra característica que possa comprometer a informação obtida.</p> <p>A descrição deve ser suficiente para permitir que outros determinem a adequação da informação para responder às questões da avaliação.</p> <p>Documentação e descrição insuficientes das fontes de informações podem reduzir a credibilidade da avaliação.</p>	<p>A4 - Fontes de Informações Defensáveis As fontes de informações utilizadas no estudo foram descritas? Os procedimentos com base nos quais se definiu a amostra foram descritos? Os processos de coleta de dados foram documentados? Cópias dos instrumentos de coleta de dados foram incluídas em anexo ao relatório de avaliação? As alterações ocorridas na amostra selecionada foram documentadas? O impacto das alterações sobre os resultados foi avaliado? Os critérios utilizados para encerrar a coleta de informações nos casos em que as informações foram coletadas mediante processo interativo foram descritos e justificados?</p>		

(continua)

(continuação)

Orientações	Questões	Escala	Comen- tário
<p>Os procedimentos para obtenção das informações devem ser escolhidos ou desenvolvidos, bem como aplicados, de forma que assegure que a interpretação dos dados seja válida para os fins da utilização pretendida.</p> <p>Validade diz respeito à fidedignidade e confiabilidade das inferências que são feitas para os resultados do processo de coleta de informações.</p> <p>Processos: testes padronizados, testes escritos, questionários, avaliação baseada em desempenho, observações, entrevistas, análise de documentos (dentre outros).</p> <p>Validação é o processo de compilação das evidências que suportam as interpretações e usos dos dados e informações coletadas, utilizando-se um ou mais desses instrumentos e procedimento.</p>	<p>A5 - Informações Válidas</p> <p>Os procedimentos utilizados na avaliação foram confrontados com os objetivos do programa para se mensurar o grau de convergência entre eles? Quem fez isso?</p> <p>As justificativas para seleção dos procedimentos e as evidências que suportam a utilização da metodologia proposta estão descritas no relatório?</p> <p>Foram utilizados procedimentos novos que demandaram interpretação cautelosa?</p> <p>Foram utilizados múltiplos procedimentos para se obter uma avaliação mais compreensiva, mas de forma parcimoniosa?</p>		
<p>Os procedimentos de obtenção das informações devem ser escolhidos ou desenvolvidos, bem como aplicados, de forma a assegurar que as informações obtidas por meio deles sejam suficientemente consistentes para a sua utilização.</p> <p>Confiabilidade refere-se ao grau de consistência da informação obtida por meio de um processo de coleta de dados.</p> <p>Quando analisar o grau de confiabilidade é importante considerar a unidade de análise ou o grupo alvo que a avaliação está focalizando. Se a unidade de análise é um grupo, não é apropriado avaliar a análise de consistência que foca no indivíduo como unidade de análise.</p>	<p>A6 - Informações Confiáveis</p> <p>Os procedimentos para obtenção das informações foram escolhidos, desenvolvidos e aplicados, de forma a assegurar que as informações obtidas por seu intermediário fossem suficientemente consistentes para a sua utilização?</p>		
<p>As informações coletadas, processadas e relatadas na avaliação devem ser sistematicamente revisadas e qualquer erro encontrado deve ser corrigido.</p> <p>Controle sistemático de informações implica etapas a serem seguidas para assegurar que toda informação, se já existente ou recentemente coletada, está dentro do possível livre de erros e mantida em segurança.</p>	<p>A7 - Informações Sistematizadas</p> <p>Há alguma sistemática para checar os erros na coleta, processamento e relato das informações e resultados?</p> <p>Existem mecanismos de controle para assegurar a integridade das informações originais?</p> <p>Existem procedimentos para guarda e manutenção das informações coletadas?</p>		

(continua)

(continuação)

Orientações	Questões	Escala	Comen- tário
<p>As informações quantitativas de uma avaliação devem ser adequada e sistematicamente analisadas de forma a responder efetivamente às questões da avaliação.</p> <p>O método analítico escolhido deve ser apropriado para atender às questões da avaliação e as características da informação que será analisada.</p>	<p>A8 - Análise Quantitativa da Informação</p> <p>O procedimento analítico utilizado foi apropriado para a questão da avaliação e a natureza do dado?</p> <p>As limitações do procedimento analítico foram descritas?</p> <p>O impacto das limitações sobre a análise efetuada foi descrito?</p>		
<p>A informação qualitativa em uma avaliação deve ser adequada e sistematicamente analisada de forma que as questões da avaliação possam ser efetivamente respondidas.</p> <p>A análise qualitativa em um estudo avaliativo é o processo de compilamento, análise e interpretação qualitativa da informação sobre um programa em resposta a uma questão em particular.</p>	<p>A9 - Análise Qualitativa da Informação</p> <p>O procedimento analítico e o método foram apropriados para atender à questão estudada e à natureza qualitativa da informação coletada?</p> <p>As limitações do procedimento analítico foram descritas?</p> <p>A consistência dos achados foi testada?</p>		
<p>As conclusões apresentadas em uma avaliação devem ser explicitamente justificadas para que os stakeholders possam avaliá-las.</p> <p>As conclusões de uma avaliação, que representam os julgamentos e recomendações devem ser defensáveis e defendidas.</p> <p>As conclusões devem ser baseadas nas informações pertinentes coletadas, devem incorporar os resultados da análise e devem ser acompanhadas de informações completas sobre como a avaliação foi conduzida.</p>	<p>A10 - Conclusões Justificadas</p> <p>As conclusões respondem às questões requeridas pelos públicos da avaliação?</p> <p>As conclusões refletem com fidedignidade os procedimentos e achados da avaliação?</p> <p>As conclusões estão relacionadas com o relatório? V. A5</p> <p>O relatório apresenta explicações plausíveis dos achados?</p>		
<p>Os procedimentos para elaboração do relatório devem salvaguardar contra a ocorrência de distorções decorrentes de sentimentos pessoais e outras de qualquer natureza, de forma que o relatório reflita fielmente a avaliação dos achados.</p>	<p>A11 – Relatório Imparcial</p> <p>Existem procedimentos para averiguar a consistência e imparcialidade do relatório?</p>		

(continua)

(conclusão)

Orientações	Questões	Escala	Comen- tário
<p>A própria avaliação deve ser formativa e somativamente avaliada contra esses e outros padrões pertinentes, de forma a garantir que foi adequadamente conduzida e orientada e, em complemento, os stakeholders podem pertinentemente examinar suas forças e fraquezas.</p>	<p>A12 - Meta-avaliação</p> <p>Passos:</p> <p>Aloque suficiente orçamento e outros recursos para conduzir meta-avaliações formativas e somativas.</p> <p>Designe alguém responsável para documentar e avaliar os processos e produtos da avaliação de programa.</p> <p>Designação de coordenador de equipe.</p> <p>Determine e registre as regras para a equipe.</p> <p>Reserve a autoridade final para editar o relatório da meta-avaliação para os meta-avaliadores.</p> <p>Determine e registre que públicos receberão o relatório da meta-avaliação e como os relatórios serão transmitidos.</p> <p>Avalie os instrumentos, a coleta de dados, os dados disponíveis, codificando e analisando a avaliação de programa para determinar como efetivamente esses passos serão executados.</p> <p>Considere que a meta-avaliação pode estar sujeita à replicação e avaliação e mantenha os registros de todos os passos da meta-avaliação, informações e análises.</p>		