

**DIRETRIZES PARA AVALIAÇÃO DE PROJETOS E
PROGRAMAS DE INVESTIMENTO SOCIAL PRIVADO**
Estudo da aplicabilidade dos *Standards de Avaliação de Programas* definidos
pelo *The Joint Committee of Standards for Educational Evaluation* em programas
e projetos de investimento social privado.

Eduardo Marino

Dissertação apresentada ao Programa de Pós-Graduação em Administração de Empresas da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo para obtenção do título de Mestre em Administração.

Orientadora: Profa. Dra. Rosa Maria Fischer

**São Paulo
2003**

**Dedico aos meus pais
pelo empenho e
exemplo na educação dos
seus oito filhos.**

AGRADECIMENTOS

Uma das necessidades mais fortes sentidas ao término da jornada de produção desta dissertação é a de agradecer a todos os que de alguma forma contribuíram com este trabalho. Início agradecendo à Rosana, minha esposa, e a meus filhos Ian, Thomas e Ângelo pela aceitação de minha ausência. À Rosana agradeço as palavras de incentivo, as sucessivas leituras, críticas e contribuições.

À Professora Rosa Maria Fischer, minha orientadora, agradeço pela confiança depositada na condução de um estudo sobre um tema ainda pouco difundido no Brasil nas Escolas de Administração, pela orientação sobre os melhores caminhos a percorrer e pela disponibilização de apoio logístico do CEATS.

Quero fazer um agradecimento nominal aos entrevistados que disponibilizaram tempo e compartilharam seus conhecimentos em vários momentos das pesquisas. Avaliadores: Dra. Ana Maria Saul, Dra. Maria do Carmo Brant, Dra. Marialva Tavares, Dr. Oswaldo Tanaka, Dra. Thereza Penna Firme. Gerentes: Andréia Buoro, Beatriz Cardoso, Celso Santiago, Diomar Silveira e Inês Kisil Miskalo. O agradecimento é extensivo às organizações que gentilmente cederam suas experiências em avaliação: Fundação Abrinq, Fundação Telefônica, Fundação Vale do Rio Doce, Instituto Ayrton Senna e Raytheon Corporation.

Muitas das condições para a viabilização deste trabalho foram proporcionadas por Vitae – Apoio à Cultura, Educação e Promoção Social. Agradeço à Sra. Regina Weinberg e ao Dr. Getúlio Carvalho pelo apoio e flexibilidade no uso do tempo, que tornaram possível conciliar minhas atribuições profissionais com as demandas de estudo e pesquisa.

Agradeço ao Marcelo Nagamatsu Tsutsui e Andréia Santos pelos trabalhos de transcrição das entrevistas. Conteí nesta dissertação com análise atenta e precisa nos trabalhos de revisão de Lia Trzmielina, a quem sou muito grato.

Faço agora menção da minha gratidão aos amigos com os quais vivenciei experiências de trabalho e de estudo e que agregam as dimensões do entusiasmo e ideais no campo da avaliação e do desenvolvimento de pessoas e organizações: Antonio Luis, Daniel Brandão, Flora Lovato, Marina Magalhães, Margareth Goldemberg, Thomaz Chianca e Rogério Silva.

Algumas instituições e pessoas foram cruciais para o despertar do meu interesse pelo tema da avaliação e pela possibilidade de vislumbrar um trabalho acadêmico neste campo. Quero agradecer à Fundação W. K. Kellogg, e aos Drs. Marcos Kisil e Francisco Tancredi que, em 1995, tornaram viável minha participação no *Kellogg Latin America Evaluation Seminar* promovido pela *Western Michigan University*.

SUMÁRIO

1 INTRODUÇÃO AO ESTUDO.....	9
1.1 JUSTIFICATIVA.....	9
1.2 OBJETIVOS.....	17
1.3 PROBLEMA DE PESQUISA.....	18
1.4 DEFINIÇÕES DOS PRINCIPAIS CONCEITOS UTILIZADOS NA DISSERTAÇÃO.....	21
2 CONSTRUÇÃO DO REFERENCIAL TEÓRICO.....	23
2.1 INVESTIMENTO SOCIAL CORPORATIVO.....	23
2.1.1 Da filantropia tradicional para ações empresariais: a emergência do investimento social corporativo.....	23
2.1.2 A ação social das empresas no Brasil - a face filantrópica.....	28
2.1.3 O investimento social corporativo no Brasil – As parcerias e ações integradas.....	32
2.2 EVOLUÇÃO DA AVALIAÇÃO.....	35
2.2.1 A primeira geração: Medição.....	36
2.2.2 A segunda geração: Descrição.....	38
2.2.3 A terceira geração: Julgamento.....	40
2.2.4 Problemas observados nas três primeiras gerações.....	42
2.3 MODELOS DE AVALIAÇÃO CONTEMPORÂNEOS.....	44
2.3.1 Modelos <i>versus</i> Abordagens <i>versus</i> Tipos de avaliação.....	44
2.3.2 Classificações e agrupamentos das abordagens de avaliação.....	45
2.3.3 Abordagens de avaliação para o século XXI.....	50
2.3.4 Abordagens de avaliação da agenda social e <i>advocacy</i> para o século XXI.....	55
3 BASE TEÓRICA UTILIZADA.....	63
3.1 OS <i>STANDARDS</i> ENQUANTO PARADIGMA.....	63
3.2 APRESENTAÇÃO DOS 30 <i>STANDARDS</i> DE AVALIAÇÃO DE PROGRAMAS.....	67
3.2.1 <i>Standards</i> de Utilidade.....	67
3.2.2 <i>Standards</i> de Viabilidade.....	71
3.2.3 <i>Standards</i> de Propriedade.....	72

3.2.4	<i>Standards</i> de Precisão.....	74
3.3	PESQUISAS SOBRE A APLICABILIDADE INTERNACIONAL DOS <i>STANDARDS</i>	77
4	MODELAGEM DA PESQUISA.....	79
4.1	MÉTODO ADOTADO.....	79
4.2	SELEÇÃO DOS CASOS.....	82
4.3	DETALHAMENTO DA PRIMEIRA ETAPA DA PESQUISA.....	83
4.3.1	Análise e interpretação dos dados da primeira etapa.....	86
4.4	DETALHAMENTO DA SEGUNDA ETAPA DA PESQUISA.....	88
4.4.1	Análise e interpretação dos dados da segunda etapa.....	89
5	ANÁLISE E DISCUSSÃO.....	90
5.1	ANÁLISE DESCRITIVA.....	90
5.1.1	Características das instituições.....	90
5.1.2	Características dos programas e das avaliações realizadas.....	91
5.1.3	Perfil dos Entrevistados.....	103
5.2	ANÁLISE DE RESULTADOS - PRIMEIRA ETAPA DA PESQUISA - META-AVALIAÇÃO.....	105
5.2.1	Desempenho geral dos casos em relação aos <i>Standards</i> ...	106
5.2.2	Análise de performance dos casos segundo proposta de Stufflebeam	113
5.2.3	<i>Standards</i> críticos.....	117
5.3	SEGUNDA ETAPA DA PESQUISA - POSICIONAMENTO EM RELAÇÃO AOS <i>STANDARDS</i>	119
5.3.1	Uso e utilidade dos <i>Standards</i>	120
5.3.2	Posicionamento dos entrevistados em relação aos <i>Standards</i>	122
5.3.3	Análise cruzada de casos e recomendações quanto aos <i>Standards</i> com menor grau de concordância.....	125
6	CONCLUSÕES E RECOMENDAÇÕES.....	146
6.1	LIMITAÇÕES DA PESQUISA.....	152
7	ANEXOS.....	154
ANEXO A	- INSTRUMENTO PARA META-AVALIAÇÃO.....	154
ANEXO B	- MEMÓRIA DE CÁLCULO DA ANÁLISE DO CONJUNTO DE CASOS PELA PROPOSTA DE STUFFLEBEAM.....	165
ANEXO C	- ROTEIRO DE ENTREVISTAS – GERENTES.....	167
ANEXO D	- ROTEIRO DE ENTREVISTAS – AVALIADORES.....	176
ANEXO E	- QUESTIONÁRIOS DE CONCORDÂNCIA.....	187
8	REFERÊNCIAS BIBLIOGRÁFICAS.....	193

LISTA DE FIGURAS

Figura 1	Distribuição das seis abordagens avaliativas de acordo com as dimensões Utilitária – intuitivas/ pluralísticas.....	49
Figura 2	Classificação das melhores abordagens da avaliação segundo análise baseada nos <i>Standards</i> de Avaliação de Programas.....	54
Figura 3	Método de estudo de casos múltiplos adaptado.....	81
Figura 4	Desempenho individual dos casos segundo análise de Stufflebeam.....	116

LISTA DE TABELAS

Tabela 1	Descrição resumida de 22 abordagens de avaliação divididas nas 4 categorias.....	51
Tabela 2	Principais usos para a palavra <i>Standards</i>.....	65
Tabela 3	Exemplo de um <i>Standard</i> e seus <i>checkpoints</i> específicos.....	84
Tabela 4	Resumo das características das organizações participantes da pesquisa.....	90
Tabela 5	Desempenho geral dos casos nos <i>Standards</i>.....	112
Tabela 6	Performance do conjunto de casos nos <i>Standards</i>.....	114
Tabela 7	<i>Standards</i> Críticos.....	118
Tabela 8	Percentuais de concordância com relação aos <i>Standards</i>.....	123
Tabela 9	Grau de concordância/ discordância dos <i>Standards</i>.....	124
Tabela 10	Resumo das análises e recomendações.....	145

RESUMO

Esta dissertação busca contribuir com a crescente demanda por parâmetros para o planejamento e a implementação de avaliação de projetos e programas sociais operados por organizações do terceiro setor com recursos advindos de doadores do setor privado no Brasil. Em muitos países, avaliadores têm-se organizado em associações profissionais, buscando fomentar pesquisas, discussões sobre *best practices* e definição de códigos de ética. Nos Estados Unidos, representantes da *American Evaluation Association* e de outras associações locais desenvolveram um conjunto de 30 *Standards* baseados em práticas generalizadamente aceitas para orientar o planejamento e para avaliar a qualidade de avaliações realizadas. Procurou-se verificar junto a um grupo de avaliadores e gerentes a aplicabilidade desses *Standards* no Brasil. Foram conduzidos cinco estudos de caso, considerando-se como “casos” processos de avaliação realizados por avaliadores externos. A maior parte dos 30 *Standards* foram aceitos como aplicáveis tal qual foram originalmente propostos. Alguns deles receberam sugestões de modificação em seus textos para melhor adequação ao contexto brasileiro. O estudo revelou que há no país um ambiente propício à sua utilização. Revelou também que o atributo de maior fragilidade dos casos estudados é o da “Utilidade”. A pesquisa aponta desafios e questões, como a participação de interessados e a necessidade de distinção entre as práticas de pesquisa e avaliação, para tornar mais efetivas as avaliações realizadas nas organizações.

Palavras-chave: avaliação, projetos, programas, terceiro setor, *Standards*.

ABSTRACT

This essay aims at contributing to the increasing demand for standards to guide the planning and implementation of evaluation of social projects and programs carried out by third sector organizations with resources provided by private sector donors in Brazil. In many countries, evaluators are getting organized into professional associations, aiming at fostering research, debates on best practices and definition of professional responsibility codes. In the United States of America, representatives of the American Evaluation Association and of other local societies developed a group of 30 standards based on generally accepted practices to orient planning and evaluation of the quality of the evaluations carried out. This study tried to verify with a group of evaluators and managers the applicability of these standards in Brazil. Five case studies were conducted, considering “cases” evaluation processes led by external evaluators. Most of the 30 standards were accepted as originally proposed. Some changes were suggested to some of them in order to better adapt them to the Brazilian reality. The study showed a favorable environment for the utilization of the standards in the Country and, according to its results, “Utility” is the most fragile attribute of the studied cases. The research points out challenges and questions, such as the participation of the involved people and the necessity of distinguishing research and evaluation practices, in order to increase the effectiveness of the evaluations developed by the organizations.

Key-words: Standards, evaluation, projects, programs, third sector.

1 INTRODUÇÃO AO ESTUDO

1.1 JUSTIFICATIVA

A terminologia “terceiro setor” surgiu no início da década de 70, a partir de uma discussão entre pesquisadores da Administração Pública norte-americana sobre as polarizações entre o público e o privado. Etzione (1973) apud Ruckle (1993), em artigo da *Public Administration Review*, fez menção a um movimento que vinha ocorrendo de forma silenciosa e que não se situava na polarização entre Estado e mercado, que na época ocupavam o foco central dos debates. Esse novo ator situava-se entre os dois pólos e, na verdade, representava um terceiro setor que talvez viesse a ser a mais importante alternativa para as décadas seguintes, não tomando o lugar de qualquer um dos outros dois, mas aproximando e balanceando os papéis de ambos.

No Brasil, a aceitação deste grupo de organizações como um “setor”, bem como o interesse acadêmico por sua caracterização e funcionamento, só ocorreram em meados dos anos 90 (Falconer, 1999).

Em 1992, Salamon & Anheier definiram as organizações que compõem o terceiro setor de acordo com seis características: 1. As organizações são formalmente constituídas; 2. Estão localizadas fora do aparato formal do Estado; 3. Não

distribuem lucros entre seus sócios; 4. São autogovernadas; 5. Possuem em seu quadro um corpo de voluntários; e, 6. Sua Atuação voltada para o benefício público.

Esta pesquisa revelou dados surpreendentes. O número de organizações do terceiro setor nos Estados Unidos era próximo de um milhão e essas organizações geravam cerca de 7,5 milhões empregos. Dados de 1997 do *Foundation Center* apontavam a existência de 1,2 milhões de Organizações Sem Fins Lucrativos, sendo que 40 mil destas eram Fundações e o setor arregimentava o trabalho de 11% da população economicamente ativa.

Em outra pesquisa de Salamon & Anheier (1994), realizada para procurar conhecer as dimensões das organizações sem fins lucrativos em outros 12 países, entre eles o Brasil, constatou-se um papel significativo do setor na geração de empregos. Nos países considerados desenvolvidos, o número de pessoas empregadas pelas organizações do terceiro setor, em 1990, era: na Inglaterra 950 mil, na França 800 mil, na Alemanha um milhão e no Japão 1,4 milhão. Verificou-se que países como França, Alemanha e Inglaterra, que possuem um sistema de bem-estar social mais fortemente patrocinado pelo governo do que os Estados Unidos, também possuem um setor filantrópico. O que significa dizer que o fato da filantropia ser bem desenvolvida nos Estados Unidos, ou em outros países, não está relacionado somente a uma falta de investimento do governo no bem-estar social. Existe uma mobilização por parte da sociedade civil que independe da ação governamental nas questões relacionadas a saúde, educação, cultura e assistência social.

Infelizmente, na pesquisa, não foi possível levantar números sobre empregos no Brasil. Entretanto, em pesquisa posterior, Landim (1998) concluiu que havia cerca de 220 mil organizações sem fins lucrativos no Brasil, que empregavam cerca de 2% da população ativa do país.

Interagem com o terceiro setor instituições privadas e públicas, sendo que as organizações do setor que oferecem serviços nas áreas de educação, cultura, assistência social e saúde são definidas por Fernandes (1994) como organizações privadas com finalidade pública. Os recursos que viabilizam a operação dessas organizações são provenientes de diferentes fontes (indivíduos, empresas, fundações, governo, agências de cooperação internacional e outros), sendo que, embora haja estudos que procurem determinar a dimensão e o escopo das ações dessas organizações, pouco se sabe sobre o impacto dessas ações. Um estudo realizado por Riddell et al. (1997) apud Roche (2000), após análise de 240 relatórios de projetos desenvolvidos em 29 países, com complementação da pesquisa por meio de 13 estudos de caso, constatou que era difícil avaliar com precisão o verdadeiro impacto do trabalho de organizações não governamentais – ONGs, e que havia pouco consenso sobre quais ferramentas e métodos eram os mais apropriados para realizar essa verificação.

O texto abaixo, veiculado pela imprensa americana, reforça os achados acima:

“Além de relatos informais ou de estudos acadêmicos minuciosos ocorridos vários anos após o fato, muitas organizações não governamentais não tinham a menor idéia dos efeitos que os programas desenvolvidos por elas estavam tendo nas pessoas e nas comunidades servidas por elas.”¹

A situação citada, que dizia respeito ao cenário vivido por ONGs americanas há poucos anos atrás pode, com segurança, ser extrapolada para a realidade do terceiro setor brasileiro nos dias atuais. São poucas as organizações do terceiro setor, tanto as financiadoras de projetos quanto as que desenvolvem atividades diretas na comunidade,

¹ Tradução de parte de artigo de J. EVANS publicado no jornal *The Washington Post* de 31 de Janeiro de 1997, p. 11. apud *Empowerment Evaluation and Foundations : A Matter of Perspectives*.

que possuem mecanismos sistemáticos que permitem a produção e a divulgação de informações relevantes e confiáveis sobre os resultados alcançados em seus programas, projetos e demais atividades.

Este fato torna-se particularmente preocupante quando se considera o grande volume de recursos investidos por empresas, agências internacionais e indivíduos, bem como o crescente espaço que o terceiro setor ocupa na mídia em geral. Em um futuro não muito distante, os profissionais da mídia não mais se contentarão em registrar e divulgar projetos bem intencionados que “pareçam” produzir efeitos positivos na redução dos problemas sociais que procuram solucionar. Nem tampouco os membros dos conselhos das ONGs, ou das empresas que investem em programas de atuação social ficarão satisfeitos com opiniões baseadas, quase exclusivamente, em aferições subjetivas como forma de prestação de contas do resultado produzido com o investimento realizado.

A necessidade de adoção de práticas sistemáticas de avaliação de projetos e programas sociais já vem sendo apontada, há algum tempo, por aqueles que investem nas organizações do terceiro setor. Em estudo realizado em 1996 com 183 empresas (que apóiam ações sociais) de um *ranking* de 1000 companhias nos EUA (*Fortune 1000 companies*), 36% mencionaram a adoção de práticas de avaliação como de alta prioridade, 36% como de moderada prioridade e 25% como de baixa prioridade. Nenhuma delas, contudo, afirmou realizar sistematicamente avaliação de seus programas sociais. Em 1998, a mesma pergunta foi feita para o total dos 35 membros associados naquele época do GIFE – Grupo de Institutos, Fundações e Empresas – localizado em São Paulo, que congrega hoje 63 instituições que fazem investimento em projetos e programas sociais. Quarenta por cento dos respondentes elegeram a adoção de práticas de avaliação como de alta prioridade. Apenas um dos

respondentes afirmou realizar avaliações sistemáticas de seus programas (Marino, 1998).

Pesquisa realizada pelo IPEA – Instituto de Pesquisas Econômicas Aplicadas para verificar a atuação social das empresas brasileiras constatou que, na região sudeste, 300 mil empresas aplicaram cerca de R\$ 3,5 bilhões em ações sociais no ano de 1998. Apenas 12% declararam fazer algum tipo de avaliação documentada sobre o investimento. Desse universo, em 60% dos casos a avaliação foi realizada pela própria empresa e em 40% pela instituição beneficiada. Não houve portanto, naquele ano, avaliações externas por instituições ou consultores especializados (IPEA, 2000). Em pesquisa sobre o perfil de atuação de 48 respondentes do “Censo GIFE”, foi constatado que 91,7 % deles realizam avaliações ao final de projetos. O estudo não permite aferir, no entanto, o tipo de metodologia empregada nessas avaliações e o grau de precisão que elas obtêm (GIFE, 2001).

Nos Estados Unidos, na década de 70, foram fundadas a *American Evaluations Research* e a *The Evaluation Network*, que em 1986 fundiram-se para a criação da *American Evaluation Association*. Esta instituição, que atualmente congrega por volta de 1400 membros, participou da formulação dos “*Standards for Evaluations of Educational Programs, Projects and Materials*” (*Standards*), a partir de um comitê formado em 1981 por membros que naquele período pertenciam às duas instituições. O processo de criação dos *Standards* levou cinco anos e em 1986 foi publicada a primeira edição que, posteriormente, em 1994, foi revisada e ampliada, visando a contemplar outros campos além do educacional.

Em muitos países, nos últimos cinco anos, percebe-se um forte movimento de estruturação do campo da avaliação. Países como a Nigéria, África do Sul, Quênia, Japão, Rússia, Sri Lanka, Coreia, Israel e Itália, entre outros, organizaram, neste

período, associações, sociedades e redes que congregam os profissionais da área de avaliação, juntando-se assim ao grupo de nações que já há alguns anos vem estruturando os procedimentos avaliativos, tais como a Austrália, Nova Zelândia, Canadá, EUA, Inglaterra e França. Em países latino-americanos tem-se notícia de esforços desta natureza na Costa Rica, onde se estabeleceu uma associação nacional de avaliação². No período em que os países supracitados reuniram-se em mesa-redonda, não havia registros de nenhum movimento dessa natureza no Brasil.

Recentemente foi iniciado no Brasil um movimento para a constituição da Rede Brasileira de Avaliação – RBA. A RBA é formada por profissionais e instituições que atuam no campo da avaliação ou tem interesse no desenvolvimento de conhecimentos e práticas neste campo de conhecimentos. Possui grupos em quatro estados (Bahia, Recife, Rio de Janeiro e São Paulo) e no Distrito Federal. Os grupos foram constituídos por profissionais provenientes de organizações do terceiro setor e do setor público, como: UNICEF, UNESCO, Care Brasil, Fundação Abrinq, Universidade de São Paulo, Universidade Federal do Rio de Janeiro e Ministério do Planejamento. Em recente exercício de planejamento estratégico, representantes dos cinco grupos definiram como missão da RBA: “Desenvolver e fortalecer cultura e práticas de avaliação de interesse público no Brasil, por meio da articulação entre pessoas e organizações, da gestão do conhecimento, do aprimoramento profissional e da defesa da avaliação como ação política transformadora.”³

O quadro de emergência e consolidação do terceiro setor como um importante agente do desenvolvimento social no Brasil, aliado à necessidade de geração de conhecimento para fortalecer a ação das empresas e das organizações da sociedade

² Mesa-redonda *American Evaluation Association Conference*, 2000.

³ Resultado do *Primeiro Seminário de Planejamento Estratégico da Rede Brasileira de Avaliação*, ocorrido em São Paulo, em 15 e 16 de julho de 2003.

civil, justifica pesquisas nas várias áreas de conhecimento abrangidas pela atuação desse setor. O estudo da temática da avaliação faz-se urgente, no entanto, por tratar de questões como a ética, tomada de decisão e aprendizagem. A ética está relacionada com a necessidade de prestar contas do uso dos recursos privados para fins públicos. A tomada de decisão diz respeito à necessidade do setor superar a fase da ação com base em informações subjetivas ou com base em opiniões, e passar a considerar em suas decisões informações obtidas por meio de metodologias apropriadas. A idéia da aprendizagem pela prática da avaliação é recente e está relacionada à possibilidade dos indivíduos aprenderem sobre os processos que envolvem seu trabalho, suas relações com outros membros da organização, e sobre si próprio e sua relação com a cultura da organização. (Preskill & Torres, 2000).

A inquietação e a movimentação de profissionais interessados no desenvolvimento do conhecimento sobre avaliação de programas é também um fator que justifica o esforço da presente pesquisa, de trazer para o Brasil o debate sobre princípios e diretrizes que orientam as melhores práticas avaliativas nos Estados Unidos e que já foram discutidos em países da Europa e Ásia.

Na primeira parte desta monografia são apresentadas uma introdução com informações contextuais e argumentações que justificam o estudo realizado. São também delineados os objetivos, o problema de pesquisa e as principais definições dos conceitos utilizados.

Na segunda parte, buscou-se identificar as características do investimento social privado no Brasil, relacionando alguns aspectos contextuais com proposições de estágios de desenvolvimento das práticas de investimento social de empresas. Em seguida, é apresentado um histórico do desempenho da metodologia da avaliação de programas e projetos sociais, a análise de 22 modelos de avaliação e, em detalhes,

quatro modelos contemporâneos de avaliação voltados ao campo social. Os últimos dois tópicos da segunda parte fornecem a base de compreensão e fundamentação do embasamento teórico que é utilizado na terceira parte. Nela é destacada a idéia dos *Standards*, os quais se constituem no paradigma da avaliação de programas e projetos vigente nos Estados Unidos. São apresentados também estudos sobre a aplicabilidade dos *Standards* em outros países.

Na parte quatro discorre-se sobre a metodologia utilizada para se verificar a aplicabilidade dos *Standards* no Brasil e, na quinta parte são apresentadas as análises dos dados, as discussões e recomendações.

Na última parte do documento são apresentadas as conclusões e limitações deste estudo.

1.2 OBJETIVOS

O principal objetivo da pesquisa é verificar, junto a um grupo de profissionais e pesquisadores que realizam avaliação de programas no Brasil, a aplicabilidade dos *Standards* definidos pelo *The Joint Committee of Standards for Educational Evaluation*, levando-se em consideração o estado da arte da avaliação de projetos e programas sociais das empresas que fazem investimento social privado no Brasil. O estudo procurou verificar, também, se a adoção destes *Standards* pode vir a atuar como impulsionadora de boas práticas de avaliação de programas no Brasil.

A pesquisa foi operacionalizada através de delineamento e análise das práticas atualmente adotadas pelas organizações que se utilizam de metodologias mais sofisticadas de avaliação, a partir dos parâmetros sugeridos pelos *Standards*. Essa opção metodológica permitiu levar a cabo um outro objetivo da pesquisa que é a definição de eventuais usos dos *Standards* pelas empresas e organizações brasileiras que realizam investimento social privado.

Finalmente pretende-se, a partir do estudo, disponibilizar uma reflexão sobre parâmetros e procedimentos para profissionais e pesquisadores interessados no estudo da utilidade, viabilidade, propriedade e precisão de avaliações.

1.3 PROBLEMA DE PESQUISA

O fenômeno da emergência do terceiro setor como segmento co-responsável pelo desenvolvimento social e as decorrentes alianças intersetoriais entre empresas, organizações da sociedade civil e organizações governamentais, tem gerado novos desafios em termos de competências pessoais e organizacionais para esses três atores. As empresas que empreendem ações de investimento social deparam-se com a necessidade de mensurar resultados e gerar informações em campos não tão precisos quanto o financeiro ou o contábil. As organizações da sociedade civil são pressionadas a provar a efetividade de suas ações. Neste contexto, pesquisadores, consultores e gestores das organizações que investem recursos e operam projetos e programas buscam modelos e parâmetros para proceder à sua avaliação.

Existe no país algum conhecimento acadêmico acumulado sobre avaliação de políticas públicas nas áreas de educação e saúde que é recente, já que esta prática foi retomada no país na década de 80, durante o processo de redemocratização. Os modelos de avaliação em larga escala de políticas públicas nem sempre são adequados a avaliações de projetos e programas do terceiro setor. As dimensões e interesses do setor com projetos e programas são distintos. Os projetos e programas são de menor dimensão, e visam sobretudo a geração de modelos de intervenção para serem adotados como políticas públicas. Há, portanto, a necessidade de avaliações que cumpram o papel de *accountability* (confiabilidade da relação fiduciária) mas, sobretudo, que indiquem a possibilidade de disseminação de iniciativas bem sucedidas.

Essas considerações preliminares sustentam a formulação do problema de pesquisa delineado na seguinte indagação:

“Os *Standards* de Avaliação de Programas definidos pelo *The Joint Committee of Standards for Educational Evaluation* podem servir de parâmetros orientadores à avaliação de programas e projetos apoiados por organizações brasileiras que realizam investimento social privado?”

Além disso, pretendeu-se elaborar possíveis adaptações oriundas da resposta a esta questão através da proposição de um rol de *Standards* que atendam as especificidades de instituições semelhantes às que foram objeto deste estudo.

As questões abaixo foram propostas como perguntas-diretrizes para a pesquisa:

1. Em que medida as avaliações externas de programas e projetos de organizações do terceiro setor obedecem aos *Standards*?
2. Como os pesquisadores e consultores se posicionam em relação aos *Standards*?
3. Quais dos *Standards* são condizentes e aplicáveis, considerando-se o atual estado da arte da avaliação de programas e projetos sociais desenvolvidos por organizações que fazem investimento social privado no Brasil?
4. Quais *Standards* podem ser aceitos, quais devem ser rejeitados e quais podem ser propostos pelo grupo de participantes da pesquisa?

A pesquisa possui caráter exploratório, descritivo e qualitativo por atender a duas intenções: 1) Conhecer em maior profundidade as práticas de avaliação de projetos e programas em um setor específico, e 2) apresentar e discutir

características específicas de processos avaliativos de um conjunto de programas. Essas intenções enquadram-se na caracterização de pesquisa exploratória proposta por Selltiz et al. (1985). A característica qualitativa do estudo refere-se à escolha do método de estudo de casos múltiplos. A necessidade de compreender com maior profundidade a relação formal estabelecida entre avaliadores externos e clientes e as escolhas metodológicas feitas para a implementação de avaliações externas orientaram, por sua vez, a escolha dessa metodologia.

1.4 DEFINIÇÕES DOS PRINCIPAIS CONCEITOS UTILIZADOS NA DISSERTAÇÃO

Para melhor delimitação da abrangência deste estudo, são apresentadas abaixo as definições dos principais conceitos nele utilizados. As quatro primeiras, sugeridas pelo comitê formulador dos *Standards*, são úteis para a compreensão da amplitude desses termos quando utilizados neste documento. A quinta definição esclarece o conceito de investimento social privado, entendida como o mais adequado para caracterizar a natureza das ações sociais estudadas.

Avaliação: identificação, esclarecimento e aplicação de critérios, passíveis de serem defendidos publicamente, para determinar o valor (mérito e relevância), qualidade, utilidade, efetividade ou importância de determinado objeto – programa, projeto, material educacional e outros – avaliado em relação aos critérios estabelecidos.⁴

Programas: atividades educacionais que são oferecidas em bases contínuas. Exemplos de atividades contempladas em programas: estímulo à leitura em escolas, de treinamento em empresas ou em setores militares, programas educacionais de saúde ou programas educacionais contínuos.⁵

⁴ De acordo com “*The Program Evaluation Standards*” (*The Joint Committee of Standards for Educational Evaluation*, 1994).

⁵ Idem

Projetos: atividades educacionais que são oferecidas por um período determinado de tempo. Exemplos de atividades: *workshop* de curta duração de caráter instrucional, projetos piloto de desenvolvimento social ou projeto de desenvolvimento de metodologias para gerenciamento de pessoas. Projetos que são institucionalizados tornam-se programas.⁶

Standards para Avaliação: princípio mutuamente acordado por pessoas engajadas em uma prática profissional, que, se for satisfeito, poderá assegurar a qualidade ou credibilidade da prática profissional da avaliação.” Optou-se pela não tradução deste termo pelo fato de não ter sido encontrada uma palavra em Português capaz de transmitir o significado deste termo tal qual ele é proposto pelo Comitê formulador. As palavras *critérios*, *padrões*, *parâmetros* ou *princípios* comumente utilizadas para a tradução do termo, não se mostram adequadas conforme pode ser melhor observado na parte deste estudo onde se discute a base teórica utilizada.⁷

Investimento Social Privado: uso planejado, monitorado e voluntário de recursos privados – provenientes de pessoas físicas ou jurídicas – em projetos de interesse público. Incluem-se neste universo as ações sociais protagonizadas por empresas, fundações ou institutos de origem empresarial ou constituídas por famílias ou indivíduos.⁸

⁶ De acordo com “*The Program Evaluation Standards*” (*The Joint Committee of Standards for Educational Evaluation*, 1994).

⁷ Idem.

⁸ GIFE, 2002

2 CONSTRUÇÃO DO REFERENCIAL TEÓRICO

2.1 INVESTIMENTO SOCIAL CORPORATIVO

2.1.1 Da filantropia tradicional para ações sociais empresariais: a emergência do investimento social corporativo

O esforço para aclarar e delinear o papel das empresas no atual contexto globalizado foi iniciado na década de 80 e vem-se consolidando. Já é amplamente aceita a idéia de que faz-se necessária a atuação social das empresas para que no longo prazo os consumidores de baixa renda ampliem seu poder de compra. O Estado mínimo, a forte ação da sociedade civil e a consciência do consumidor impelem as empresas a uma definição mais acurada de seu papel no desenvolvimento social.

Um indicador destas mudanças na forma das empresas realizarem suas funções sociais é o surgimento de entidades como o BSR – *Business for Social Responsibility*, criado em 1993 nos Estados Unidos, que congrega atualmente empresas norte-americanas e transnacionais que movimentam cerca de 3 trilhões de dólares. O BSR posiciona-se como uma organização global que apóia empresas associadas para serem bem sucedidas nas suas práticas de respeito a valores éticos, a

pessoas, comunidades e meio ambiente. Possui parceiros no Reino Unido, Japão, Israel e Brasil, que atuam buscando propósitos semelhantes.⁹

Da mesma forma que seu parceiro e inspirador BSR, foi criado no Brasil, em 1998, o Instituto Ethos de Responsabilidade Social buscando influenciar e oferecer ferramentas de gestão que visam à melhoria da qualidade das relações da empresa com seus colaboradores, clientes e fornecedores, com a comunidade e com o poder público. O conjunto de 765 empresas filiadas ao Instituto Ethos emprega cerca de 1 milhão de pessoas e soma 28% do PIB brasileiro.¹⁰

O número de empresas associadas às diferentes instituições que se ocupam da discussão e disseminação dos conceitos e das práticas da responsabilidade social é significativo; entretanto, isso não assegura uma clareza conceitual sobre o termo em si e as várias dimensões compreendidas nesta idéia. As terminologias **responsabilidade social**, **cidadania corporativa** e **investimento social corporativo** muitas vezes são utilizadas como sinônimo ou como termos complementares, como se observa na colocação de Falconer (1999, p.16):

“Cidadania Empresarial é um termo que tem sido utilizado para descrever o papel de responsabilidade social e ambiental das empresas. Não se trata exclusivamente de filantropia, no sentido da caridade desinteressada, mas de investimento estratégico: um comportamento de aparência altruísta, como a doação a uma organização sem fins lucrativos, que atende a interesses (mesmo que indiretos) da empresa, como a contribuição à formação de uma imagem institucional positiva ou o fortalecimento de mercados consumidores futuros”

As dimensões tratadas pelas organizações que defendem e disseminam a idéia da responsabilidade social incluem aspectos que vão além do investimento em

⁹ Fonte: *Business for Social Responsibility*. Disponível em < www.bsr.org >. Acesso em 05 de maio de 2003.

¹⁰ Fonte : Instituto Ethos. São Paulo. Disponível em < www.ethos.org.br >. Acesso em 05 de maio de 2003.

projetos ou programas ambientais e sociais, abrangendo questões relacionadas à gestão da empresa como, por exemplo, a preocupação com valores éticos e com a relação com clientes e fornecedores. A terminologia **investimento social corporativo** apresenta-se como mais apropriada para delinear os investimentos que empresas realizam em projetos e programas sociais externos, de implementação própria ou em parceria com organizações da sociedade civil. Apesar da precisão desta terminologia, no que se refere à delimitação da ação social das empresas, as práticas neste campo ainda encontravam-se em fase de estruturação no final da década de 90, conforme destacou Weeden (1998, p.1):

“Todo ano empresas desembolsam bilhões de dólares em algo definido como ‘relações externas’, e elas fazem isto sem investir os mesmos esforços que fazem quando utilizam-se de parâmetros gerenciais para acompanhar seus negócios. ‘Stakeholders’ encontram-se muito confusos sobre o real valor da miscelânea da função ‘soft’ da empresa nas atividades de relações com a comunidade, responsabilidade social ou filantropia corporativa.”

O autor observou que no final da década de 90 houve queda no investimento em ações sociais pelas empresas norte-americanas em função dos processos de reestruturação organizacional pelos quais elas passavam. Neste período de intensa revisão nos processos, redução de custos e enxugamento de pessoal, as empresas voltaram-se para si próprias, preocupando-se menos com suas relações externas, resgatando sua função primordial de maximizar seus lucros. O contexto de otimização de recursos levou à revisão dos propósitos e estratégias da então chamada **filantropia corporativa**. A falta de clareza sobre os reais resultados obtidos pelas práticas filantrópicas, aliada ao ambiente de escassez de recursos, contribuiu para a emergência da idéia de que o investimento social faria sentido naquele contexto se ajudasse os negócios a serem melhor sucedidos.

Fischer (2002) chamou atenção para a necessidade de distinguir as ações sociais orientadas apenas por motivações mercadológicas das ações que de fato exprimem uma preocupação com o desenvolvimento social. A autora propõe dois tipos de ação social das empresas para identificar as diferentes formas de atuação social: 1) A empresa considera as ações de investimento social como um empreendimento capaz de gerar resultados para a comunidade ou para determinada população-alvo; e, 2) A empresa realiza as ações sociais na ótica negocial, orientada para obter retorno de imagem na marca e no desempenho de determinado produto.

Austin (2001) também vai além da idéia de que o investimento social corporativo deve estar preponderantemente relacionado ao sucesso nos negócios. O autor sugere três estágios na relação de colaboração entre as empresas e as organizações sem fins lucrativos. No primeiro estágio, definido como **filantrópico**, a relação é unidirecional, referindo-se às práticas filantrópicas corporativas tradicionais. Doa-se sem se preocupar com retorno para a empresa ou mesmo sem a preocupação em acompanhar os resultados da ação social em si. O segundo estágio é o **transacional**, quando é estabelecida uma relação de troca na qual são explícitas as expectativas de resultados para ambos os lados. Inclui-se nesse estágio de relação o *marketing* de causa e o patrocínio a projetos sociais que geram visibilidade para a empresa doadora. No terceiro estágio, denominado **integrativo**, há um encontro entre as missões da empresa e das outras organizações envolvidas na parceria, gerando ações de interesse mútuo, afetando simultaneamente as concepções e práticas dos envolvidos. Para o autor, a atuação social da empresa em cada um desses estágios está fortemente determinada pelos padrões da cultura organizacional. Em cada estágio do *continuum* da colaboração há um acréscimo de complexidade nos sistemas de gerenciamento dessas relações.

As idéias de Austin alinham-se a uma perspectiva ampla da proposta de **cidadania corporativa** proposta por Logan, Roy & Regelbrugge (1997), que une os interesses de negócios da empresa com a perspectiva de promoção do bem comum no longo prazo. O modelo propõe quatro níveis de interesse da empresa: 1) Interesses comerciais; 2) Benefícios imediatos a partir do investimento social para melhorar a imagem; 3) Interesses comerciais expandidos; e, 4) Promoção do bem comum.

As diferentes visões sobre a atuação social das empresas sugerem que o investimento social corporativo pautar-se pela busca de uma sociedade mais justa, atendendo aos interesses de diferentes parceiros que interagem em projetos e programas destinados às mais diversas necessidades sociais e ambientais. O sucesso dessas iniciativas estará vinculado à utilização de metodologias de planejamento, monitoramento e avaliação adequadas ao tipo de ação desenvolvida e ao grau de complexidade do contexto e das articulações necessárias para atingir os propósitos desejados.

2.1.2 A ação social das empresas no Brasil – a face filantrópica

O IPEA realizou entre 1999 e 2002 uma pesquisa nacional visando a conhecer diferentes aspectos da ação social das empresas brasileiras. A pesquisa informou que 462 mil empresas, ou seja, 59% das 782 mil empresas do País declararam realizar, voluntariamente, ações sociais para a comunidade. O volume de recursos aplicados no período de um ano foi de R\$ 4,7 bilhões, equivalentes a 0,4% do PIB do país no ano de 2000. O estudo verificou que 39% das empresas declararam que pretendia ampliar os investimentos em ações sociais nos próximos anos.

A pesquisa do IPEA adotou como conceito para definir *ação social empresarial* o seguinte:

*“Qualquer atividade que as empresas realizam para atender às comunidades, nas áreas de assistência social, alimentação, saúde, educação, meio ambiente e desenvolvimento comunitário, dentre outras. Essas atividades abrangem desde pequenas doações eventuais a pessoas ou instituições até grandes projetos mais estruturados, podendo, inclusive, estender-se aos empregados da empresa e a seus familiares.”*¹¹

As informações sobre a região sudeste, que concentra mais da metade do setor produtivo do País, possibilitam diagnosticar o estágio de estruturação da ação social das empresas. A pesquisa revelou as seguintes informações sobre essa região:

¹¹ Peliano, 2000a, p.6

- O montante investido pelas 300 mil empresas no período de 1 ano foi de R\$ 3,5 bilhões;
- 95 mil empresas apoiaram ações exclusivamente para a comunidade;
- 78 mil desenvolvem ações somente para seus funcionários¹²;
- Cerca de 200 mil empresas desenvolveram ações mistas, contemplando a comunidade e os funcionários;
- 61% das microempresas (de 1 a 10 empregados) da região sudeste fizeram algum tipo de contribuição;
- 90% das empresas com mais de 500 empregados declararam ter apoiado ações para a comunidade;
- 57% apoiaram ações assistenciais e 40% atenderam as comunidades por meio de ajuda alimentar;
- 14% apoiaram ações educacionais, sendo que 43% das grandes empresas investiram nessa área enquanto apenas 9% das pequenas estiveram envolvidas com este tipo de investimento;
- Em 86% das pequenas empresas, os donos declararam-se responsáveis pela ação social. Nas grandes empresas, 72% das ações sociais são conduzidas pelos diretores;
- 81% dos empresários atendem as comunidades por motivos humanitários;
- 2% das empresas declararam comunicar sua atuação social.

Estes números revelam a atuação significativa das empresas para além da função de gerar empregos e lucros. A preocupação com os funcionários e com a

¹² Excluiu-se nas *ações aos funcionários* os benefícios obrigatórios por lei como: contribuição ao Sistema “S” (SEBRAE, SESI, SESC, SENAI, SENAC E SENAR), normas trabalhistas (periculosidade, vale-transporte, salário família, etc.).

comunidade não está restrita apenas às grandes empresas; mais de 150 mil pequenas empresas realizaram ações sociais no ano em que foi realizada a pesquisa. As empresas com mais de 500 funcionários ocupam um papel de destaque no campo social, já que a grande maioria (90%) das organizações deste porte fez algum tipo de investimento e quase metade delas investem em ações educacionais (43%).

Os aspectos relacionados à motivação e ao *modus operandi* das atividades sociais levantados na pesquisa sugerem um estágio ainda pouco estruturado em termos de profissionalização da gestão técnica e administrativa dessa atuação empresarial. Na maioria das pequenas (86%) e grandes empresas (72%) constatou-se que as decisões e a condução das ações são realizadas por proprietários e dirigentes e não por profissionais especializados nessas atividades.

À análise dessas informações, somam-se as seguintes constatações: a maior motivação é humanitária (81%), apenas 12% realizam algum tipo de acompanhamento e ainda há pouca preocupação em se comunicar resultados (2% comunicam). Desta forma, pode-se depreender que, além do baixo nível de profissionalização, não há necessariamente um interesse ligado ao negócio quando a empresa opta por realizar alguma ação social, como expressam os pesquisadores do IPEA:

“A reduzida participação das empresas em atividades de educação, alfabetização e qualificação profissional, associada à prioridade conferida às áreas de assistência e alimentação, parecem indicar que a atuação do setor privado em atividades comunitárias não obedece a uma lógica de interesse estritamente mercantil, mas sim ao atendimento difuso das principais carências sociais que afetam os pobres brasileiros. Ou seja, o movimento em prol dos mais necessitados é muito mais de cunho assistencial do que um investimento na melhoria da qualidade de uma mão-de-obra potencial.”¹³

¹³ Peliano, 2000b, p.16

A análise acima sugere que a relação de colaboração entre as empresas e as organizações sociais da região mais desenvolvida do Brasil encontram-se ainda em grande medida no estágio que Austin define como **filantrópico**. Essas ações são baseadas na relação doador/donatário, limitam-se ao atendimento de um pedido de uma instituição social ou de indivíduos e o acompanhamento no uso dos recursos é praticamente inexistente, o que sugere pouca clareza sobre o resultado para ambas as partes. As instituições recebem recursos para manutenção de seus serviços ou para pequenas melhorias em suas instalações físicas, enquanto que as empresas não recebem benefícios, além da satisfação de seus proprietários ou diretores, por apoiarem uma boa obra social.

Há, entretanto, no universo das grandes empresas, aquelas que atuam de forma profissionalizada, bastante orientadas para resultados sociais aliados aos do negócio. Essas empresas vêm desenvolvendo desde a década de 80 um percurso de construção de práticas de uma ação social sistemática e profissionalizada, orientadas para a busca de resultados sociais e empresariais. O estudo do IPEA identificou ações sociais melhor estruturadas em empresas com mais de 500 funcionários, sendo que algumas delas criam institutos e fundações especificamente para tratarem dessas ações.

2.1.3 O investimento social corporativo no Brasil – As parcerias e ações integradas

A idéia de investimento social corporativo como prática sistemática e estruturada foi introduzida no Brasil pela ação de empresas transnacionais (na maioria norte-americanas) que trouxeram de seus países as práticas das boas relações com a comunidade. No princípio da década de 80 iniciou-se um movimento liderado pela Câmara Americana de Comércio-AMCHAM, visando à disseminação das práticas de investimento de empresas em programas e projetos sociais e ambientais. A AMCHAM instituiu em 1982 o Prêmio ECO, com o objetivo de premiar ações exemplares e projetos de atendimento a necessidades de caráter público patrocinados ou implementados por empresas. Desde sua criação, o Prêmio já contou com a participação de 1285 empresas, premiou 111 delas e a soma dos investimentos realizados nos projetos inscritos somam valores na ordem de US\$ 2,6 bilhões.¹⁴ As ações da AMCHAM foram potencializadas no final da década de 80 com a criação do GIFE – Grupo de Institutos, Fundações e Empresas, a partir de um comitê formado ao final de um seminário organizado pela AMCHAM. Faziam parte deste comitê empresas como IBM, American Express, Bradesco e Dow Química. Somaram-se a essas empresas gestores da Fundação W. K. Kellogg, Instituto C&A e Vitae - Apoio à Cultura, Educação e Promoção Social, instituições com larga experiência no apoio a projetos nas áreas de educação, cultura e desenvolvimento

¹⁴ Câmara Americana de Comércio. São Paulo. Disponível em < www.amacham.org.br >. Acesso em 12 de maio de 2003.

social e. Após visitar centros de estudos e associações de fundações e empresas nos Estados Unidos e realizar uma série de discussões sobre as práticas de investimento social privado no Brasil, o grupo constituiu-se formalmente como uma associação. O GIFE tem como missão: “Aperfeiçoar e difundir conceitos e práticas do uso de recursos privados para fins públicos.”¹⁵

O GIFE faz uma distinção entre investimento social corporativo e investimento social privado, por contemplar entre seus associados fundações e institutos criados por famílias. Define investimento social privado como “uso planejado, monitorado e voluntário de recursos privados – provenientes de pessoas físicas e jurídicas – em projetos de interesse público. Incluem-se no universo do investimento social privado as ações sociais protagonizadas por empresas, fundações e institutos de origem empresarial ou instituídos por famílias ou indivíduos.”¹⁶

Em pesquisa realizada no ano de 2000 junto aos então 48 associados, constatou-se um investimento de R\$ 437,09 milhões. As informações sobre as práticas gerenciais revelam um elevado nível de institucionalização: 58,3 % das entidades associadas estão estruturados como fundações de direito privado:

- 31,3% estão estruturadas como associações sem fins lucrativos;
- 73% fazem planejamento estratégico;
- 87,5% fazem monitoramento de projetos;
- 91,7% fazem avaliação de resultados de projetos;

Educação é o principal foco de investimento (85,4%) dos associados do GIFE, seguido da área de saúde (43,7%), e cidadania (41,7%), que envolve questões

¹⁵ Fonte: Investimento Social Privado no Brasil – Perfil dos Associados do GIFE. São Paulo, 2001.

¹⁶ GIFE, 2001, p.11

relacionadas aos direitos das crianças e dos adolescentes, desigualdade e exclusão social desse público. O interesse preponderante do investimento em educação revela-se superior ao do grupo pesquisado pelo IPEA (14%), talvez porque estas entidades de origem empresarial comunguem uma visão de que é alto o impacto que projetos nesta área podem trazer para o público beneficiado e para aqueles que praticam as ações sociais de forma planejada e estruturada.

É possível depreender, a partir das informações citadas, que o conjunto de organizações ligadas ao GIFE possui um elevado grau de profissionalização. A observação da preocupação com o planejamento estratégico e com o monitoramento e avaliação de projetos sugere maior probabilidade de ações colaborativas voltadas a atender interesses mútuos nas relações de parceria.

2.2 EVOLUÇÃO DA AVALIAÇÃO

Embora nos últimos 30 anos a temática da avaliação tenha despertado grande interesse em parte de pesquisadores norte-americanos, este campo ainda era considerado por este mesmo grupo como uma área em fase de maturação.

“A avaliação como um campo de estudo estabelecido está agora vivenciando a fase final da adolescência. A fase de rápido crescimento da infância deu-se no final da década de 60 e início da década de 70. A partir de meados da década de 70 este campo de estudo tornou-se mais seguro de si próprio, introspectivo, marcando a sua fase de pré-adolescência. Atualmente, nos anos 80, a avaliação está vivendo a transição de final da adolescência para a fase adulta”.¹⁷

A declaração acima é aceita parcialmente pelos autores que a citam, que propõem que ao rótulo “adulto” seja acrescido “jovem”, sendo então mais aceitável o estágio de “jovem adulto”. Para o alcance da maturidade plena haveria ainda uma boa distância a ser percorrida, já que apesar do rápido crescimento observado como campo profissional em maturação, há ainda que se discutir se a avaliação é uma disciplina ou uma transdisciplina, como propõe Scriven (1991).

Guba & Lincoln (1989) descreveram o processo de amadurecimento da avaliação, definindo os diferentes sentidos e usos atribuídos a esta metodologia, relacionados ao contexto histórico e pressupostos filosóficos utilizados por teóricos, pesquisadores e praticantes ao longo dos anos. Os autores propõem três gerações de desenvolvimento, apresentadas a seguir.

¹⁷ Conner, Altman & Jackson, 1984, p.13 apud Worthen & Sanders, 1988

2.2.1 A primeira geração: Medição

A primeira geração é caracterizada como o período em que houve grande esforço para o desenvolvimento de testes utilizados no ambiente escolar e que foram posteriormente utilizados para fins militares e empresariais.

Os autores descrevem o desenvolvimento de testes na França onde, no início do século XX, o psicólogo Alfred Binet tentou utilizar técnicas psicométricas para identificar crianças com retardamento mental que, segundo seus professores, precisavam ser excluídas das salas de aula, já que atrapalhavam o rendimento das crianças “normais”. Binet avançou em seus estudos e desenvolveu um teste que determinava a “idade mental” de crianças. Em 1916, Louis Terman revisou o teste de Binet, transformando-o no teste de QI, que passou a ser largamente utilizado no sistema de ensino dos Estados Unidos. Neste período, foi implementada a primeira Agência de Pesquisa na cidade de Nova York, com o objetivo de realizar pesquisas no sistema escolar local, utilizando as “novas técnicas de medição”.

Os testes de QI foram posteriormente utilizados na seleção do pessoal do Exército durante a Primeira Guerra Mundial, por meio de um instrumento desenvolvido pela Associação Americana de Psicologia (APA). A utilização dos testes de QI, neste período, coincide com a ascensão das ciências sociais e com o desabrochar da administração científica, conforme os comentários de Guba & Lincoln (1989, p.5,6):

“Um segundo fator contextual a estimular o processo de testagem era a emergência do movimento da administração científica no mundo dos negócios e na indústria. Se o ser humano é o elemento mais importante na produção de bens e serviços, a função do gerente é tornar seu trabalho o mais efetivo e eficiente possível.”

Estes autores identificam o papel do avaliador neste período como essencialmente técnico, cabendo a ele conhecer a variedade de instrumentos disponíveis, a fim de avaliar a variável escolhida para a medição.

Worthen & Sanders (1987) também reconhecem o período entre 1900 e 1930 como o período do “movimento dos testes”. Os autores descrevem o mesmo percurso proposto por Guba & Lincoln, no qual a avaliação é realizada por meio de testes aplicados em crianças no sistema escolar. Posteriormente outros testes são utilizados na seleção de soldados para a Primeira Guerra. Os autores sinalizam a mudança de paradigma no final dos anos 30, com a adoção por algumas escolas de currículos progressistas baseados nos escritos de John Dewey¹⁸.

¹⁸ Filósofo da Educação das décadas de 30 a 50, precursor da aprendizagem com base na experiência (*experiential learning*).

2.2.2 A segunda geração: Descrição

De acordo com Guba & Lincoln, a segunda geração da avaliação surgiu da necessidade de superar a principal deficiência das avaliações de primeira geração, que tinha como objeto de avaliação apenas os indivíduos.

Após a Primeira Guerra, as escolas secundárias americanas começaram a receber alunos que reingressavam no sistema escolar após o período da guerra, sendo que muitos deles não haviam sequer complementado o período elementar antes dela. O sistema vigente exigia a aprovação por meio de um teste padronizado para atender um currículo unificado denominado *Carnegie*. O teste era aplicado aos alunos que já estavam no sistema escolar. O rigor do teste, aliado ao longo período fora do sistema escolar, dificultava o acesso aos ex-soldados.

Com a intenção de questionar a validade desta posição, Ralph W. Tyler, da Universidade de Ohio, liderou a partir de 1933 um estudo de oito anos com a utilização de currículos mais flexíveis, sem necessariamente atender aos padrões requeridos pelo *Carnegie*. Tyler e um grupo de especialistas desenvolveram testes que mediriam a aprendizagem dos alunos, de acordo com o que os seus professores tinham intenção que aprendessem. Os resultados de aprendizagem foram chamados de *objetivos*. Nascia neste momento a *avaliação de programas*.

“Enquanto as escolas secundárias que participavam do estudo planejavam seus currículos, Tyler coletou informações sobre a extensão da realização, por parte dos alunos do estudo, de seus objetivos definidos. Esta informação, juntamente com a análise dos

pontos fortes e fracos que se tornaram, dessa forma, aparentes, foi utilizada para guiar ajustes finos e revisões.”¹⁹

A chamada segunda geração da avaliação é caracterizada pela descrição de padrões e pontos fracos e fortes em relação a certos objetivos declarados. O papel do avaliador era o de *descrever* determinado processo. A partir desse momento, a medição não era mais reconhecida como avaliação e sim, como uma das ferramentas a serem utilizadas para este fim, conforme relato do próprio Tyler em uma entrevista concedida a Nowakowski em 1981 (p. 8):

“O simples desempenho de alunos em testes representa aquilo que os alunos lembram dos seus livros texto e não necessariamente o resultado do que lhe foi ensinado. Devido ao termo Teste estar relacionado a uma coletânea de itens memorizados, eu sugeri o termo “avaliação” para investigar o que de fato os alunos haviam aprendido.”

¹⁹ Guba & Lincoln, 1989, p. 9

2.2.3 A terceira geração: Julgamento

A abordagem descritiva mostrou-se deficiente na resolução de determinados problemas do sistema educacional americano, bem como em outras áreas. Iniciava-se o período da exploração espacial e a competição entre americanos e russos pressionava o desenvolvimento em várias áreas de conhecimento. Pelo fato de a abordagem descritiva estar fortemente ligada à necessidade da existência de objetivos previamente formulados, muitos dos profissionais envolvidos com projetos espaciais não se sentiam confortáveis em se comprometerem com objetivos antes de terem boa noção dos resultados para os quais as ações desenvolvidas apontavam. Além disso, entendiam que objetivos pré-estabelecidos tolhiam a criatividade.

Segundo Guba & Lincoln, o modelo desenvolvido por Tyler negligenciou o que Stake chamou de “a outra natureza” ou a “outra face” da avaliação: o **juízo**. De acordo com os autores, “O apelo para incluir o elemento julgamento no ato da avaliação marca a emergência da terceira geração da avaliação, uma geração na qual a avaliação foi caracterizada pelos seus esforços visando chegar a um julgamento, a um veredicto, e na qual o avaliador assumiu o papel de juiz, mantendo as já estabelecidas funções técnica e descritiva também.” (Guba & Lincoln, 1989, p.30)

Uma série de modelos foi criada a partir de 1967, contemplando a perspectiva do julgamento. Os modelos mais conhecidos são: 1. O CIPP – avaliação de Contexto, Insumos, Processo e Produto, que forneciam bases para guiar o desenvolvimento e implementação com êxito de um programa, e também para relatar

com credibilidade sua execução e resultados (Stufflebeam et al., 2000), e 2. O *Goal Free Evaluation Model*, orientado para avaliar efeitos resultantes de programas independentemente dos objetivos propostos (Scriven, 1996).

Scriven procurou combater a idéia da avaliação enquanto monitoramento de cumprimento ou não de objetivos com a proposição do *Goal Free Evaluation Model*. O autor sugeriu com o modelo que a avaliação deveria cumprir a função de verificação do que realmente foi realizado em termos de benefícios para o público-alvo do programa, ao invés de avaliar os objetivos que sugeririam o que o programa se propunha a fazer pela população. O mérito do programa seria estabelecido pela relação entre seus efeitos e as necessidades do público-alvo. Para o autor, não há total incompatibilidade entre os modelos orientados por objetivos e *goal free*; o que possibilitaria a utilização das duas abordagens para avaliar determinado programa seria a condução das duas avaliações por avaliadores distintos (Scriven, 1996).

2.2.4 Problemas observados nas três primeiras gerações

Guba & Lincoln (1989) apontam três deficiências nas três primeiras gerações: a tendência ao “gerencialismo”²⁰, a dificuldade em acomodar o pluralismo de valores; e, o compromisso excessivo com o paradigma científico de investigação. Essas três deficiências são detalhadas a seguir:

1) A tendência ao gerencialismo: São denominados *gerentes* todos os indivíduos que encomendam ou financiam as avaliações. Esses clientes ou patrocinadores definem junto aos avaliadores contratados seus interesses de acordo com seus parâmetros e limites que, via de regra, o isentam de prestar contas quanto aos achados produzidos ou não pela avaliação, pelo fato desta responsabilidade ser estritamente do avaliador. Um outro aspecto importante é que, cabendo única e exclusivamente ao gerente a definição sobre as questões a serem perseguidas pela avaliação, os parâmetros com os quais as informações serão julgadas e a forma como os resultados serão disseminados menospreza os outros interessados (*stakeholders*) envolvidos nos programas. Esses indivíduos talvez tenham outras questões a serem investigadas, bem como outros parâmetros de análise a sugerir.

2) A dificuldade de acomodar o pluralismo de valores: O senso comum sugere muitas vezes que as sociedades compartilham valores. A crença dos americanos de que o país transformou-se em um “grande caldeirão de culturas” e de

²⁰ Gerencialismo (de *Managerialism*) - Termo não encontrado nos dicionários consultados. Expressão aparentemente proposta pelo autor para atribuir sentido ao gerenciamento tendencioso e éticamente questionável.

que todos os imigrantes percebem-se como participantes de um só país e, portanto, possuem a mesma identidade cultural, não é válida. A discussão sobre o *juízo* em processos avaliativos surgiu no momento em que se questionava fortemente nos Estados Unidos as questões raciais e de gênero. Naquele momento, os avaliadores começaram a questionar *quem* faria o juízo e com que *objetivos*.

“Mas foi fácil fazer vista grossa ao fato de que até o desenvolvimento de um instrumento ‘objetivo’ envolve juízo de valores, ou de que o delineamento dos objetivos implica num acordo de valores, enquanto a questão da diferença de valores não foi levantada. Uma vez levantada, não podia ser colocada de volta em sua caixinha. A questão sobre os valores de quem iria predominar numa avaliação, ou, como alternativa, como é que diferenças de valores poderiam ser negociadas, aparecem agora como o problema maior a ser enfrentado.”²¹

3) O compromisso excessivo com o paradigma científico de investigação:

Durante a vigência das três primeiras gerações da avaliação, pesquisadores e praticantes utilizavam premissas do método científico no *design* de avaliações, levando o processo avaliativo a uma rigidez excessiva, fazendo com que variáveis contextuais fossem deixadas de lado, em função da busca do controle de fenômenos em desenhos experimentais, ou por meio de manipulação estatística.

²¹ Guba & Lincoln, 1989, p. 17

2.3 MODELOS DE AVALIAÇÃO CONTEMPORÂNEOS

2.3.1 Modelos *versus* Abordagens *versus* Tipos de avaliação

Os diferentes agrupamentos, classificações ou categorizações de avaliações congregam diferentes abordagens, tipos ou modelos de avaliação. Tem havido ao longo dos anos uma preocupação de vários autores que se ocupam dessa tarefa de discutir e prover distinções das formas e métodos avaliativos, e também de determinar o que são modelos, abordagens e tipos de avaliação. Stufflebeam (2001) entende que o termo *approach* (abordagem) em lugar de “modelo” é mais apropriado para cobrir a diversidade de alternativas de avaliações existentes. Segundo o autor, o termo “modelo” é tido como muito restritivo para denominar a abrangência de suas propostas avaliativas. Em relação à diferenciação entre modelo e tipo, Patton (1982) comenta que “é difícil fazer uma diferenciação entre modelos ou tipos. A distinção não é absoluta... diferentes rótulos e palavras significam diferentes coisas para diferentes avaliadores”. Dada a complexidade e diversidade desta área de conhecimento, parece ser mais sensato adotar a terminologia “abordagens” pois se refere a diferentes formas de avaliação praticadas.

2.3.2 Classificações e agrupamentos das abordagens de avaliação

O estudo do histórico da avaliação, assim como de outras áreas das ciências sociais, nos permite verificar o surgimento de novas abordagens, modelos e estratégias para compreender, intervir e mensurar mudanças na sociedade. Neste campo, na medida em que os diferentes autores foram ampliando e aprofundando seu estudo e o do papel do avaliador, surgiram novas concepções, métodos e procedimentos para praticar a avaliação de programas. Cabe neste Capítulo fazer menção à contribuição de Kuhn sobre o estudo da evolução da ciência, traçando-se um paralelo entre sua visão dos motivos pelo qual a ciência evolui com a razão pela qual as abordagens de avaliação foram surgindo e se modificando ao longo do tempo. Segundo este autor, o que diferenciou as várias escolas científicas não foi o insucesso de determinados métodos, mas o que ele chamou de “incomensurabilidade de suas maneiras de ver o mundo e de nele praticar ciência”²². Portanto, para além de um mero processo cumulativo de conhecimento, as novas abordagens de avaliação foram surgindo para atender as demandas mais complexas da sociedade.

Vários autores propuseram agrupamentos de abordagens de avaliação utilizando diferentes critérios para este fim. A Academia Nacional de Ciências dos Estados Unidos apud Patton (1982) sugere seis categorias de avaliação, segundo critérios de propósito dos estudos avaliativos: 1. Avaliação de necessidades, 2. Pesquisa básica, 3. Teste em pequena escala, 4. Avaliação de campo, 5. Análise de políticas, 6. *Accountability* fiscal, 7. *Accountability* de cobertura, 8. Avaliação de

²² Kuhn, 2000, p. 23

impacto, e 9. Análises econômicas. O autor também cita a classificação proposta pelo Comitê de *Standards* da Sociedade de Pesquisa dos Estados Unidos, definidos a partir do critério *tipos de atividades que se pretende avaliar*²³:

1. Análise de início (pré-instalação, contexto, análise de viabilidade): Essa avaliação ocorre antes do início do programa e proporciona informações para orientar seu planejamento e implementação.
2. Análise de avaliabilidade: Visa a avaliar a possibilidade de se utilizar as diferentes abordagens e métodos de avaliação quando se define o *design* de determinado processo avaliativo.
3. Avaliação formativa (de desenvolvimento, de processo): Visa a prover avaliação para a melhoria ou modificação do programa quando este está sendo implementado.
4. Avaliação de impacto (somativa, resultado, efetividade): Visa a determinar os resultados e efeitos do programa com o propósito principal de orientar decisões relacionadas à continuação, expansão e redução de atividades e de financiamento.
5. Monitoramento de programa: Patton (1982) utiliza a definição literal proposta pela ERS para esta categoria – “ esta é a menos conhecida mas, a mais utilizada.... Esta atividade envolve uma variedade de procedimentos que vão desde a verificação contínua da realização de atividades até a entrega de serviços previstos”.

²³ ERS – Evaluation Research Society, 1980, p. 45 apud Patton, 1982

6. Avaliação de avaliação (avaliação secundária, meta-avaliação): Esta categoria inclui críticas de profissionais para relatórios de avaliações, re-análise de informações e revisões externas de avaliações internas.

House (1980, 1983) apud Worthen, Sanders & Fitzpatrick (1997) propôs, a partir do estudo de correntes filosóficas do conhecimento, duas categorias de abordagens de avaliações: com base no objetivismo ou positivismo e com base no subjetivismo. As abordagens com base no objetivismo são aquelas “cientificamente objetivistas”, nas quais o avaliador deve adotar procedimentos na avaliação que gerem achados que possam ser encontrados por outro avaliador com preparo técnico similar que utilize os mesmos procedimentos. As abordagens ancoradas no subjetivismo sugerem um processo avaliativo que ocorre segundo as percepções do avaliador, estruturado na sua formação e qualificação. Neste caso, uma avaliação jamais poderá ser reproduzida por outro avaliador e gerar os mesmos resultados.

House sugere, a partir dessas duas categorias, uma outra distinção pautada na determinação de juízo de valor quando se avalia determinado programa. O autor denominou de “avaliações utilitárias” aquelas que visam ao impacto de um programa em sua amplitude, ou seja, a abrangência dos resultados de um programa na sua população-alvo; segundo suas próprias palavras, o utilitarismo refere-se à idéia de “maximizar a felicidade na sociedade” (apud Worthen, Sanders & Fitzpatrick, 1997, p. 66). Nas avaliações utilitárias o avaliador é o responsável pelo julgamento de mérito do programa.

Outra denominação sugerida para um outro conjunto de abordagens foi “avaliações intuitivas e pluralistas”. Para essas, o juízo dos resultados de um programa está baseado no impacto gerado em cada indivíduo beneficiário. Esta

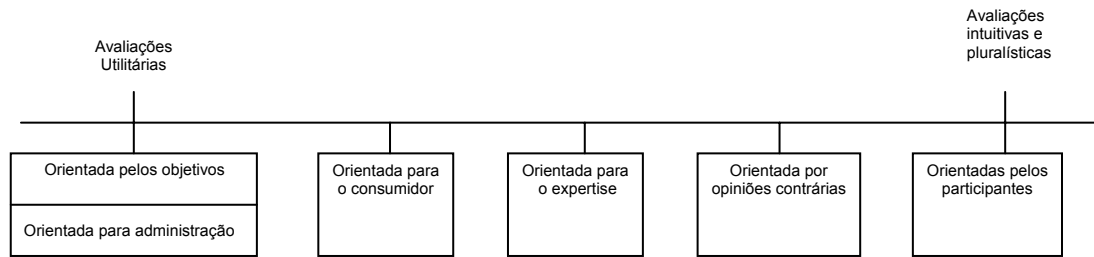
abordagem pressupõe que um programa oferece diferentes benefícios para atender a diferentes necessidades individuais; portanto, para se verificar a extensão dos resultados, faz-se necessário o julgamento individual sobre os benefícios a ele oportunizados.

Worthen, Sanders & Fitzpatrick (1997) propuseram, com base na classificação proposta por House e outros autores que já envidaram esforços em propor categorizações de abordagens, um conjunto de seis categorias que congregam um conjunto de abordagens entendidas pelos autores como as mais usuais:

1. Abordagens orientadas pelos objetivos: estão focadas em objetivos pré-determinados e mensuram a extensão na qual eles foram alcançados.
2. Abordagens orientadas para administração: a principal preocupação é identificar informações para a tomada de decisões administrativas.
3. Abordagens orientadas para consumidores: a questão central é o desenvolvimento de informações avaliativas sobre produtos (definição ampla) para uso de consumidores, em comparação a outros produtos ou serviços competidores.
4. Abordagens orientadas pelo *expertise*: dependente de julgamento da qualidade de um objeto avaliado por um especialista.
5. Abordagens orientadas por opiniões contrárias: são construídas com base em opiniões contrárias de diferentes avaliadores (pró e contra) sobre o foco central do objeto avaliado.
6. Abordagem orientada pelos participantes: os participantes (*stakeholders*) são os principais atores a determinar os valores, critério, necessidades e análise de dados.

As seis categorias seguem uma lógica baseada nas distinções propostas por entre as abordagens “utilitárias” e “intuitivas e pluralísticas”, conforme a figura 1:

Figura 1 Distribuição das seis abordagens avaliativas de acordo com as dimensões Utilitária – intuitivas/ pluralísticas



Fonte: Worthen, Sanders & Fitzpatrick, 1997, p.79

2.3.3 Abordagens de avaliação para o século XXI

Stufflebeam (2001a) realizou um extenso estudo analítico de 22 abordagens de avaliação com base nos 30 *Standards* de Avaliação de Programas elaborados pelo *Joint Committee* (1994), que serão melhor definidos na parte 3 deste estudo. As abordagens cobrem uma abrangente gama de programas. Alguns exemplos dos Programas são: programas cíclicos ou em bases contínuas referentes a currículo escolar, selos alimentação, habitação; programas anuais de vacinação; projetos com período de duração determinado; performance de pessoas em empresas ou de sistemas de serviço estatal nas mais diversificadas áreas.

O autor aponta importantes estudos analíticos de abordagens avaliativas (Stake, 1974; Hasting, 1976; Guba, 1990; Scriven, 1991, 1994; e Madaus, Scriven & Stufflebeam, 1983, 2000), que foram fonte de consulta para organizar a análise das 22 abordagens. As abordagens foram divididas em quatro categorias. A primeira é formada por duas abordagens geradoras de informações ou achados incompletos – essa categoria é denominada Pseudo-avaliações. As outras três categorias são as seguintes: Orientadas por questões e/ou métodos (13 abordagens); Melhoria ou *Accountability* (três abordagens) e Agenda social/*advocacy*²⁴ (quatro abordagens). Na tabela 1 é apresentado um resumo das 22 abordagens.

²⁴ O termo tem o significado de “defesa de causas”.

Tabela 1 - Descrição resumida de 22 abordagens de avaliação divididas nas 4 categorias

1. Pseudo-avaliações	
1.1. Estudos Inspirados em Relações Públicas	Esta abordagem inicia-se com a intenção de utilizar os dados para convencer o público (eleitores, legisladores, clientes, etc) de que determinado programa é eficaz.
1.2. Estudos Politicamente Controlados	O acesso aos achados da avaliação são controlados por motivos políticos. Pode ser lícito ou ilícito. Será ilícito quando o avaliador ou o cliente: a) reter informações da avaliação de indivíduos que possuem direito de acesso aos achados avaliativos; b) recusar-se a cumprir acordos prévios de apresentar os resultados da avaliação; c) enviesar os resultados da avaliação ao revelar apenas parte dos achados.
2. Avaliações orientadas por questões e/ou métodos (Quase-estudos avaliativos)	
2.1 Estudos baseados em objetivos	O objetivo desta abordagem é verificar se os objetivos de determinado programa foram atingidos. Os objetivos operacionais são especificados e informações são levantadas para verificação da medida de sucesso no cumprimento desses objetivos.
2.2. Estudos de <i>Accountability</i>	Limita-se a responder questões sobre resultados relacionados a investimentos efetuados. Legisladores, administradores públicos e privados e conselheiros formulam questões avaliativas relacionadas ao uso responsável de recursos no cumprimento de atividades previstas e na conquista dos objetivos previstos.
2.3. Administração de testes (objetivos)	Aplicação de testes (via de regra em estudantes) com o objetivo de verificar, através do nível de acerto em questões de múltipla escolha, a qualidade de projetos específicos, programas ou mesmo de professores individualmente. Infere-se que os resultados positivos ou negativos são reflexos diretos da qualidade dos projetos ou dos professores.
2.4. Estudos de estimativa de agregação de valor (resultados)	Estudo baseado em aplicações sistemáticas de testes que propiciam análise hierarquizada de efeitos de programas e políticas. A ênfase desta abordagem é a verificação da adição de valor na performance de estudantes de diferentes regiões, distritos, tamanho de escolas, perfil de docentes, etc.
2.5. Testes de performance	Estudos desenvolvidos nos anos 90 buscaram suprir as limitações dos testes de múltipla escolha orientados apenas para conteúdos. Os testes de performance visam a mensurar as habilidades (escrita, fala, interpretação, análise, trabalho em equipe e solução de problemas).
2.6. Estudos experimentais	Utilizam-se experimentos controlados para determinar contrastes entre grupos em relação a determinados resultados de intervenções. São também chamados de estudos quase-experimentais por responder apenas a questões específicas de intervenções de programas, devido às exigências metodológicas para viabilizar o uso do método.
2.7 Sistemas de informações administrativas	Sistemas que suprem administradores de informações para apoiar a condução e a comunicação sobre o andamento de programas.
2.8. Análise de custo-benefício	Estudos de custo-benefício são desenvolvidos através da aplicação de diversos procedimentos quantitativos, para compreender o custo total do programa e verificar se os investimentos realizados geraram retornos em objetivos alcançados e benefícios sociais.

2. Avaliações orientadas por questões e/ou métodos (Quase-estudos avaliativos)	
2.9. Audiência para esclarecimento	Terminologia oriunda da área jurídica, trata do uso da técnica <i>role-playing</i> , na qual avaliadores desempenham o papel de promotores e advogados defendendo aspectos pró e contra, baseados em evidências coletadas por ambos. O júri pode ser formado por representantes do programa e outros interessados.
2.10. Estudo de caso	Avaliações baseadas em estudos de caso tratam de descrição em profundidade, análise e síntese de um programa em particular ou de determinado objeto.
2.11. Análise crítica por especialistas	Nasceu inspirada na arte de fazer crítica literária. Assume que determinados <i>experts</i> em determinadas áreas são capazes de realizar análise em profundidade que não podem ser feitas por outros especialistas.
2.12 Avaliação baseada em teoria	São avaliações que se utilizam de uma base teórica sobre como o programa deve funcionar ou se baseiam em teorias de programas similares. A teoria orienta o avaliador na construção de perguntas e indicadores, assumindo que essa base teórica é suficientemente consistente para estabelecer relações entre ela e o que de fato ocorreu no programa.
2.13 Estudos <i>Mixed-Methods</i>	Esta abordagem surgiu a partir da discussão sobre o uso de métodos quantitativos e qualitativos. Apesar de todas as abordagens supracitadas se utilizarem dos dois métodos, os estudos <i>mixed-methods</i> justifica-se como uma abordagem à parte porque o centro das suas preocupações é discutir as vantagens e desvantagens dos diferentes métodos bem como as possíveis combinações entre eles.






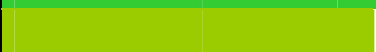
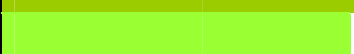


3. Abordagens orientadas para a melhoria e <i>accountability</i> de programas (abordagens avaliativas)	
3.1 Estudos orientados para <i>accountability</i> e decisões	Esta abordagem enfatiza que a avaliação de um programa deve ser utilizada de forma proativa e retroativa para ajudar a melhorar o programa e julgar o mérito e a relevância do mesmo. O que distingue este método do estudo de “sistemas de informação administrativa” é o julgamento de mérito e relevância. Nesta abordagem prevê o envolvimento de interessados na definição do foco da avaliação, para que sejam geradas informações úteis à tomada de decisão.
3.2. Estudos orientados para consumidores	Esta abordagem considera o bem estar do público-alvo (consumidor) de um programa como justificativa principal de sua implementação. Espera-se nesta modalidade de avaliação a determinação de mérito e relevância do programa em relação aos benefícios gerados aos consumidores (estudantes, professores, pais, etc.).
3.3. Abordagem de Acreditação ou Certificação	Nesta abordagem, instituições, programas e indivíduos são alvo de estudos para verificar se eles se enquadram em determinados requerimentos com vistas a assumir determinadas funções na sociedade. A acreditação ou certificação são estruturadas por padrões e critérios que acreditam ou certificam aqueles que desejam exercer função específica.

4. Abordagens voltadas à agenda social e <i>advocacy</i>	
4.1. Estudos centrados no cliente (avaliação responsiva)	Esta abordagem provém da chamada “Escola relativista da avaliação”. O avaliador atua com e para os diferentes clientes ou interessados no programa (Exemplo: administradores, educadores, legisladores, conselheiros). Diferentes e, às vezes, conflitantes valores dos interessados são considerados para julgar o programa. A diferença entre esta abordagem e a “ <i>voltada para os consumidores</i> ” está no papel do avaliador que nesse modelo assume o papel de julgar o mérito e relevância do programa, não considerando os valores dos consumidores.
4.2. Avaliação construtivista	Esta abordagem rejeita a existência de uma realidade completa e absoluta como um objeto a ser avaliado. Este tipo de estudo pressupõe que é indeterminado e mutável o conhecimento ganho a partir da construção de e interação entre indivíduos. Cabe ao avaliador assumir um consistente papel indagativo junto aos diferentes interessados, ajudando-os a se apropriarem do processo avaliativo, por meio de definição de perguntas, variáveis, estabelecimento de comunicação constante e análise e interpretação conjunta das informações.
4.3. Avaliação democrática e deliberativa	Esta abordagem é bastante recente e está estruturada em três dimensões: 1) participação democrática; 2) diálogo para credenciar interessados; e, 3) deliberação para se chegar ao julgamento de mérito e relevância de determinado programa.
4.4. Avaliação focada no uso	Esta abordagem é explicitamente orientada para assegurar que uma avaliação resulte em impacto no objeto da avaliação. A construção do estudo é feita a partir de escolhas pré-determinadas do grupo que fará uso dos resultados da avaliação, a partir das suas intenções de uso desses resultados.

Fonte: Autor

As 22 abordagens foram avaliadas utilizando-se como critério de julgamentos os 30 *Standards* de Avaliação de Programas elaborados pelo *Joint Committee* (1994). A escala de julgamento utilizada foi a seguinte: Pobre, Ruim, Bom, Muito Bom e Excelente. Para a análise das diferentes abordagens foi desenvolvido um instrumento de *Checklist* contendo 10 itens de verificação para cada um dos 30 *Standards*. O autor concluiu que 9 das 22 abordagens puderam ser considerados as melhores e as mais promissoras para serem utilizadas e melhoradas nos tempos atuais, segundo o parâmetro de julgamento adotado. Na figura 2, são apontadas as nove melhores abordagens.

Figura 2 - Classificação das melhores abordagens da avaliação segundo análise baseada nos *Standards* de Avaliação de Programas

						Pontuação total	Utilidade	Viabilidade	Propriedade	Precisão
	Pobre	Ruim	Bom	Muito Bom	Excelente					
Melhoria/Accountability										
Decisão/Accountability						92 (MB)	90 (MB)	92 (MB)	88 (MB)	98 (E)
Orientação para o Consumidor						81 (MB)	81 (MB)	75 (MB)	91 (MB)	81 (MB)
Acreditação						60 (B)	71 (MB)	58 (B)	59 (B)	50 (B)
Agenda Social/Advocacy										
Focada no uso						87 (MB)	96 (E)	92 (MB)	81 (MB)	79 (MB)
Centrada no cliente						87 (MB)	93 (E)	92 (MB)	75 (MB)	88 (MB)
Democrática e deliberativa						83 (MB)	96 (E)	92 (MB)	75 (MB)	69 (MB)
Construtivista						80 (MB)	83 (MB)	67 (B)	88 (MB)	83 (MB)
Questões/Métodos										
Estudo de caso						80 (MB)	68 (MB)	83 (MB)	78 (MB)	92 (MB)
Agregação de valor						72 (MB)	71 (MB)	92 (MB)	69 (MB)	56 (B)

Fonte: Evaluation Models – Stufflebeam, L., D., 2001

2.3.4 Abordagens de avaliação da agenda social e *advocacy* para o século XXI

Entendendo o trabalho de Stufflebeam como uma contribuição quanto à adequação e ao apontamento de tendências de abordagens de avaliação para os próximos anos, a seguir serão apresentadas as quatro abordagens indicadas como melhores para o campo da agenda social e *advocacy*.

A. Avaliação com foco na utilização (*Utilization Focused-Evaluation*)

Patton foi quem cunhou a abordagem da avaliação com foco no uso. A inclusão de expectativas de interessados no *design* de avaliações visando à utilização dos seus achados, bem como a atribuição de julgamento de valor de determinado programa aos usuários intencionais dos achados da avaliação, são as características centrais deste tipo de abordagem.

“Utilization focused-evaluation inicia-se com a premissa de que avaliações devem ser julgadas pela sua verdadeira utilidade; para tanto, os avaliadores devem facilitar o planejamento e a implementação da avaliação com cuidadosa consideração... não em abstrações mas sim levando-se em consideração as expectativas de pessoas reais no mundo real ... conseqüentemente, o foco da Utilization focused-evaluation é na intenção de uso revelada pelos usuários intencionais.”²⁵

Nesta abordagem, o avaliador deve facilitar o julgamento e o processo decisório dos principais interessados na avaliação. Não cabe ao avaliador agir a distância e proferir seu próprio julgamento como um juiz independente. Patton parte

²⁵ Patton, 1996, p.20

da premissa de que não há avaliações livres de valores. Portanto, os valores daqueles interessados no uso dos achados das avaliações devem determinar o rol de critérios de julgamento a serem utilizados para avaliar determinado programa.

Para Patton, a avaliação é muito importante para ficar sob responsabilidade apenas do avaliador. Para o autor, o papel do avaliador é atuar como um facilitador e negociador, apresentando opções aos usuários intencionais para a escolha de caminhos metodológicos para a avaliação. O autor recomenda que, quaisquer que sejam, as opções apresentadas devem atender aos *Standards* de avaliação de Programas.

B. Avaliação centrada no consumidor

Stake (1967,1975,1999) apud Stufflebeam (2001a) foi o pioneiro no desenvolvimento e aplicação das abordagens centradas no cliente, a partir de sua compreensão de que os modelos científicos de pesquisa social não se aplicam diretamente aos propósitos de uma avaliação. A abordagem centrada no cliente ou responsiva é inspirada na pesquisa-ação. A avaliação segundo esta abordagem é um processo no qual diferentes interessados no programa (financiador, equipe gerencial, público beneficiário) interagem assumindo papéis de sujeitos da avaliação, obtendo maior compreensão sobre o programa, tomando melhores decisões e proporcionando com isto melhorias nas suas ações e no programa como um todo.

As condições básicas para a utilização desta abordagem são o preparo do avaliador para atuar de maneira responsiva (orientado exclusivamente para as necessidades do cliente) e a abertura por parte do cliente para aceitar um plano de avaliação aberto, flexível e inclusivo quanto à participação de interessados. A

abordagem utiliza-se, na grande maioria de situações, de métodos qualitativos de coleta de informações. O objetivo central desta abordagem não é levantar respostas definitivas para perguntas relacionadas ao mérito e à relevância de determinado programa. Busca-se levantar em profundidade as necessidades e as expectativas dos diferentes interessados quanto ao programa e envolvê-los em todo o processo de análise e interpretação dos achados da avaliação.

C. Avaliação democrática e deliberativa

House & Howe (2000) foram os proponentes da abordagem da avaliação democrática e deliberativa. Essa abordagem segue a tendência contemporânea de avaliações que contemplam ampla participação dos interessados nas diferentes etapas do processo avaliativo. As premissas fundamentais que justificam a proposição desta abordagem são: 1) avaliações não podem ser “puras” ou isoladas do contexto sociopolítico no qual elas ocorrem; e, 2) avaliações sempre ocorrem em sistemas sociais onde existem estruturas de poder.

A avaliação democrática e deliberativa coloca em questão escolhas iniciais feitas pelo avaliador quanto à definição de foco. No processo de definição do foco, antecipam-se questões como: Quem são os interessados nos resultados da avaliação? Quais são as necessidades e questões desses interessados? Quais serão os critérios de julgamento a serem utilizados?

Essas questões suscitam uma importante indagação no campo das relações de poder: Uma avaliação a ser desenvolvida em um ambiente democrático deve atender aos interesses exclusivos daqueles que possuem maior poder financeiro (contratante)?

Para os autores, a avaliação em uma sociedade democrática deve seguir princípios de debate e buscar formas democráticas e deliberativas de julgamento. Os autores apontam três requerimentos para a avaliação democrática e deliberativa: inclusão, diálogo e deliberação. O requerimento de inclusão refere-se à consideração de interesses relevantes ao *design* da avaliação. Esses interesses não são apenas os daqueles que possuem mais poder. A inclusão deve ser compreendida segundo um princípio fundamental da democracia que é o da equidade, ou seja, os interesses relevantes não são definidos a partir do *status* social dos diferentes interessados e, sim, pela busca de um equilíbrio entre os diferentes grupos envolvidos no programa. O requerimento do diálogo trata da importância do avaliador emvidar esforços para buscar uma profunda compreensão da posição, visão e interesses dos diferentes indivíduos ou grupos interessados na avaliação. Quanto à deliberação, o entendimento desse requerimento pelos autores é de que deliberar é um “processo cognitivo, baseado em razões, evidências e princípios de argumentos válidos, que representam um importante subconjunto de cânones para a avaliação”.²⁶

Para os proponentes desta abordagem, o profissional que opta por ela não é um facilitador inocente, um filósofo ou um espectador curioso. O avaliador é um profissional consciente que adota uma conduta de reunir cuidadosamente os princípios de inclusão, diálogo e deliberação.

Os autores propõem dez perguntas que indicam aspectos a serem observados na adoção desta conduta de avaliação:

²⁶ House & Howe, 2000, p. 8

1. Quais interesses estão representados na avaliação? Os interesses e visões de todos os usuários que possuem envolvimento significativo no programa devem ser considerados.
2. Os principais interessados estão representados? Os diferentes interessados devem estar presentes ou representados no processo avaliativo.
3. Há algum interessado excluído da avaliação? Cabe ao avaliador assegurar a inclusão de grupos de menor poder (pobres, negros e outras minorias).
4. Há sérios problemas de desbalanço de poder? Muitas avaliações são controladas por interesses exclusivos de grupos com muito poder (Exemplo: clientes contratantes da avaliação).
5. Existem procedimentos para controlar o desbalanço de poder? Cabe ao avaliador criar condições para a coleta de informações, envolvendo neste processo o diálogo e a deliberação.
6. Como as pessoas participam da avaliação? Apesar do alto custo e do consumo de tempo para envolver diferentes participantes, uma séria e representativa participação de diferentes interessados no processo avaliativo pode contribuir em muito para a efetiva participação de diferentes indivíduos e grupos.
7. Quão autêntica é a participação das pessoas? Simples respostas a questionários quantitativos não podem ser consideradas como autêntica participação de interessados na avaliação. É preciso que o avaliador utilize métodos e instrumentos capazes de captar as reais percepções, opiniões e sentimentos dos respondentes ou entrevistados.
8. Quão envolvidas estão as pessoas? É necessário estabelecer critérios para um envolvimento adequado dos interessados. É preciso evitar situações extremas

como o envolvimento deles na análise estatística de dados, assim como deve-se evitar envolvimento superficial dos interessados em momentos críticos como análise de dados.

9. Existe deliberação reflexiva? A fase final de um processo avaliativo via de regra é marcada por intensa pressão de tempo para entrega do relatório final. A pressa coloca em risco todo o esforço realizado na condução da avaliação, já que se perde a oportunidade de contemplar no relatório final as reflexões de todos os interessados envolvidos nas fases anteriores.
10. A deliberação é considerada e utilizada em alguma extensão? Considera-se que quanto mais extensa for a deliberação, melhor será a qualidade dos achados. A prática da deliberação pode ajudar a sanar um problema muito comum nas avaliações que é a não coincidência das conclusões com os dados encontrados.

D. Avaliação construtivista ou naturalística

Guba & Lincoln (1989) são os autores que moldaram a abordagem construtivista ou naturalística da avaliação. Essa abordagem é caracterizada por ser uma “avaliação construtivista responsiva”. O termo “construtivista” foi sugerido para desvincular este modelo do paradigma científico; o parâmetro investigativo sob o qual a proposta se embasa é o interpretativo e hermenêutico. Quanto ao emprego do termo “responsiva”, os autores sugerem que o modelo deve ser flexível no sentido de acomodar diferentes parâmetros de julgamento e atender a diferentes expectativas, seguindo a proposta de Stake (1975) da avaliação responsiva. A abordagem

naturalística utiliza os referenciais da abordagem responsiva para definir o foco da avaliação e o modo construtivista para implementá-la.

Os procedimentos da abordagem responsiva para definir o foco de avaliação compreendem o levantamento de *reivindicações*, *preocupações* e *questões* sobre o objeto da avaliação identificadas pelos principais interessados, que são compreendidos como pessoas ou grupos que serão afetados em alguma medida pela avaliação. Uma *reivindicação* é entendida pelos autores como qualquer afirmação favorável de qualquer interessado sobre o objeto da avaliação (Exemplo: “*a afirmação de que um certo modo de ensinar a ler vai resultar em um ano de ganho nos resultados em testes padronizados de leitura*”); a *preocupação* é entendida como qualquer afirmação dos interessados que seja desfavorável ao objeto da avaliação (Exemplo: “*o ensino do uso do computador reduz a habilidade das crianças de fazer contas a mão*”); e uma *questão* é uma expressão de desconforto com assuntos polêmicos como, por exemplo, a inclusão do ensino religioso como obrigatório no ensino fundamental. Cabe ao avaliador criar condições para que os interessados expressem suas *reivindicações*, *preocupações* e *questões* e incluí-las na definição do foco da avaliação.

Guba & Lincoln (1981) apud Guba & Lincoln (1989) definem três categorias de interessados:

1. Os *agentes*: responsáveis pela concepção e implementação do objeto da avaliação, incluindo-se neste grupo os patrocinadores, consultores, equipe técnica, e equipe operacional;
2. Os *beneficiados*: beneficiários diretos ou “grupo-alvo”, que são pessoas para as quais o objeto da avaliação foi desenvolvido, e beneficiários indiretos, que

são pessoas que recebem benefícios por manterem algum tipo de contato ou relação com os beneficiários diretos. Além desses dois grupos, existem pessoas que se beneficiam com as possibilidades de serviços geradas pelo programa ou objeto (Exemplo: editores, empreiteiros, etc.);

3. As *vítimas*: pessoas afetadas por motivos de falha do objeto ou por mudanças advindas do próprio resultado da avaliação (Exemplo: jovens normais excluídos de programas para superdotados, crianças que perdem vagas em escolas em função de inclusão de cotas de minorias).

A dimensão construtivista desta abordagem está nas construções geradas pelos diferentes interessados a partir de suas *reivindicações, preocupações ou questões*. Cabe ao avaliador que optar por essa abordagem conduzir um processo de confronto das diferentes construções, processo este compreendido pelos autores como dialético hermenêutico. À medida em que as construções são confrontadas e compreendidas, algumas dessas *reivindicações, preocupações ou questões* podem ser solucionadas ou modificadas, exigindo a busca de novas informações. A partir das novas informações, novas construções são criadas e isto torna as informações mais sofisticadas. Embora procure-se um consenso, isto raramente ou nunca é possível. Para Guba & Lincoln, a iteração e reiteração da avaliação construtivista faz com que ela nunca se complete plenamente. O final da avaliação se dá por motivo de suspensão por questões de viabilidade ou por decisão dos envolvidos.

3 BASE TEÓRICA UTILIZADA

3.1 OS *STANDARDS* ENQUANTO PARADIGMA

Os *Standards* são fruto de intensos debates de representantes da comunidade científica norte-americana e surgiram da necessidade dos profissionais atuantes no campo da avaliação definirem parâmetros orientadores, tanto para a conduta de avaliadores como para a orientação de procedimentos metodológicos. Em 1974 foi formado um comitê misto indicado pela *American Educational Research Association*, pela *American Psychological Association* e pelo *National Council on Measurement in Education*, com a missão de revisar a edição de 1966 dos *Standards for Educational and Psychological Tests and Manuals* publicados pela *American Psychological Association*. Após muitos debates, o grupo chegou a uma recomendação para a criação de uma sessão específica para os *Standards tests*. Após a experiência, em 1975, as três organizações aceitaram a sugestão de desenvolver um rol de *Standards* úteis não só para testes, mas para avaliação de programas educacionais em geral. Um outro comitê foi formado e, em 1981, chegou ao produto denominado: *Standards para a Avaliação de Programas, Projetos e Materiais Educativos*.

Em 1989, em função do grande interesse pela avaliação em outros campos, um grupo mais amplo das mesmas instituições responsáveis pela criação e edição dos *Standards* para avaliação de programas educacionais formou o *The Joint Committee on Standards for Educational Evaluation* e preparou uma nova edição dos *Standards* para avaliação de programas, aplicável para as áreas de educação, saúde, militar, corporativa, governamental e serviço social. Para se chegar a um produto aceito em todo o país americano, foram envolvidas 15 entidades associativas de diversas áreas e quatro anos de trabalho foram consumidos.

O comitê chegou a um rol de 30 *Standards* que estão subdivididos sob quatro atributos básicos: utilidade, viabilidade, propriedade (*propriety*²⁷) e precisão. Os *Standards* são considerados como uma base filosófica que reflete as melhores práticas avaliativas para o período em que foram escritos. Para o comitê, deve haver um esforço para que os diferentes *Standards* para a avaliação sejam testados e analisados sobre sua aplicabilidade em diferentes realidades. O comitê aponta como possível utilidade uma orientação para o planejamento e realização de meta-avaliação. Há uma expectativa da aplicação desses *Standards* em outros países, com registro da discussão sobre sua aplicabilidade na Alemanha, Coréia e Suíça (Russon, 2000).

O percurso percorrido para a definição dos 30 *Standards* utilizados como objeto central do presente estudo remete a compreendê-los como um paradigma. Kuhn (2000, p. 218) conceitua paradigma como “constelação de crenças, valores, técnicas,...partilhadas de uma comunidade determinada”.

²⁷ De acordo com o dicionário Michaelis, traduz-se o termo *propriety* como: adequação, retidão, justeza, decoro, decência, boas maneiras. Segundo o dicionário Houaiss o termo propriedade refere-se *ao que é próprio* ou *direito legal de possuir algo*. O sentido atribuído a propriedade neste estudo está relacionado à conduta apropriada ou adequada do avaliador no respeito aos direitos humanos, leis e normas vigentes no contexto em que se realiza uma avaliação.

Segundo Morin apud Moraes (1997), um paradigma possui uma natureza lógica entre um conjunto de conceitos mestres. A reunião de práticas generalizadamente aceitas pela comunidade de pesquisadores e avaliadores fundamenta-se no avanço histórico da avaliação enquanto área de conhecimento e em acordos entre a comunidade sobre diferentes abordagens em uso.

O entendimento do sentido atribuído à palavra *Standards* pelo *The Joint Committee* difere em algumas das aplicações usualmente dadas ao termo. Singer (1996) reuniu quatro usos mais comumente dados a essa palavra:

Tabela 2 – Principais usos para a palavra *Standards*

Uso	Exemplo
Qualidade	Mínimo nível ou “base”. Exemplo: Tocar uma sonata corretamente, ou a execução soberba de uma sonata.
Padronização/ norma	Produto ou serviço idêntico em forma a outro. Frequentemente utilizado no âmbito industrial com a idéia de padronização e intercâmbio de produtos. Norma para comparação.
Ideal	Meta a ser cumprida em determinado período. Posteriormente se torna uma norma a ser superada.
Critério	Especificação ou aplicação

Fonte: Singer D. B., 1996, p.208

O *Joint Committee* define os *Standards* como “princípio mutuamente acordado por pessoas engajadas em uma prática profissional, que, se for satisfeito, poderá assegurar a qualidade ou credibilidade da prática profissional da avaliação.” (1994, p. 2)

Quanto à sua aplicação, o *Joint Committee* (1994, p. 8, 10) sugere o seguinte:

...“os Standards são princípios orientadores e não uma prescrição mecânica. Eles contêm orientações para evitar eventuais erros e se fundamentam em práticas generalizadamente aceitas, além de propor diretrizes que refletem as ‘best-practices’ em avaliação de programas atualmente.”

...“os Standards devem ser utilizados como orientação para avaliar planos e relatórios... devem ser utilizados como meio para troca de informação sobre a qualidade da avaliação entre o cliente, os avaliadores e os vários stakeholders envolvidos...”

Os formuladores dos *Standards* atribuem ao termo o sentido de qualidade, somado a um sentido de ideal ou diretriz a ser perseguida. Há também na expressão “princípio” a idéia de doutrina fundamental²⁸. A idéia dos *Standards* enquanto normas ou padrão está descartada já que não houve intenção por parte dos seus formuladores de definir um rol de padrões a serem seguidos ou reproduzidos. O entendimento do termo enquanto critério também está descartado, já que eles não representam especificações e sim diretrizes.

²⁸ Dicionário Eletrônico Michaelis. Universo On Line. São Paulo. Produzido por Amigo Mouse LTDA.

3.2 APRESENTAÇÃO DOS 30 *STANDARDS* DE AVALIAÇÃO DE PROGRAMAS ²⁹

3.2.1 *Standards* de Utilidade

Os *Standards* relacionados à “utilidade” têm o objetivo de assegurar que uma avaliação estará a serviço das necessidades de informação dos seus usuários intencionais. O atributo “utilidade” de uma avaliação foi proposto, segundo Patton (1996), como resultado de reflexões sobre as críticas às práticas adotadas das décadas 60 e 70. A maturidade dos participantes do *Joint Committee* quanto a questões como o avaliador como único a emitir juízo de valor, o atendimento de necessidades específicas dos contratantes da avaliação e a necessidade de um distanciamento da avaliação do paradigma científico, contribuiu para a proposição do atributo “utilidade”. Patton cita uma fala de Stufflebeam (1980) em uma entrevista sobre a lógica utilizada para definir a ordem em que são apresentados os *Standards* na publicação editada em 1981:

“Os Standards que serão publicados determinam que as avaliações devem atender a quatro requisitos. São eles: utilidade, viabilidade, propriedade e precisão. Penso que foi interessante a decisão do Joint Committee sobre essa ordem em particular. O raciocínio utilizado foi o de que uma avaliação não deve ser realizada se não houver um estudo prospectivo sobre o seu uso por diferentes audiências. Segundo, ela não deverá ser realizada se não for viável em termos práticos, políticos e econômicos. Terceiro, o Joint Committee entende que a avaliação não deve ser realizada se não for possível demonstrar que ela foi conduzida de maneira ética e honesta. Finalmente, uma

²⁹ Tradução para o Português de Chianca com adequações. Fonte capa e contracapa da publicação; *The Program Evaluation Standards*, 2nd Edition.

*vez considerados a utilidade, viabilidade e propriedade, pode-se tratar de aspectos relacionados aos procedimentos técnicos utilizados na avaliação*³⁰.

A utilidade como primeira instância de orientação para uma avaliação busca atender os requisitos das abordagens: *Utilization Focused Evaluation*, Centrada no Consumidor, Democrática/Deliberativa, e Construtivista. Essas abordagens possuem um elemento comum que é a participação dos interessados desde a fase de planejamento da avaliação. O atributo utilidade presta-se também para delinear a distinção entre avaliação e pesquisa. Cronbach e Suppes (1969) apud Worthen e Sanders (1987) fazem a seguinte diferenciação: *pesquisa busca conclusões enquanto avaliação leva a decisões*. Segundo os autores, o avaliador busca informações úteis aos processos decisórios de gerentes e administradores, governantes e outros. O pesquisador realiza uma investigação para chegar a conclusões, a partir de palpites ou hipóteses por ele formulados. O objetivo é entender determinado fenômeno; portanto, as informações encontradas são um meio para este fim.

Worthen e Sanders consideram que pesquisa e avaliação têm em comum a necessidade do uso de técnicas e métodos empíricos de investigação. Os autores afirmam, no entanto, que os objetivos da avaliação e da pesquisa são muitas vezes confundidos. Os critérios para julgar a qualidade de uma pesquisa podem ser úteis para desfazer eventuais dúvidas entre as duas atividades. A validade interna e externa de uma pesquisa são critérios determinantes quanto à sua precisão e generalização. Para a avaliação, não se entende a generalização como um critério prioritário; os *Standards* de precisão visam a assegurar informações confiáveis para a tomada de decisão e não necessariamente para garantir a generalização de achados.

Os sete *Standards* de Utilidade são os seguintes:

³⁰ Apud Patton, 1997, p. 16

1) Identificação dos interessados: Pessoas envolvidas ou afetadas pela avaliação devem ser identificadas, de forma que suas necessidades possam ser atendidas.

2) Credibilidade do avaliador: As pessoas que conduzem a avaliação devem ser confiáveis e competentes, a fim de que os resultados da avaliação obtenham grau elevado de credibilidade e aceitação.

3) Escopo e seleção de informações: Informações coletadas devem ser selecionadas de forma ampla para que possam abordar perguntas pertinentes sobre o programa e trazer respostas a necessidades e interesses dos clientes e de outros interessados pela avaliação.

4) Identificação de valores: As perspectivas, os procedimentos e as justificativas usados para interpretar os resultados da avaliação devem ser cuidadosamente descritos, de forma que as bases para julgamentos de valores fiquem bem claras.

5) Clareza dos relatórios: Relatórios de avaliação devem descrever claramente o programa avaliado, incluindo seu contexto e os propósitos, procedimentos e conclusões da avaliação, de forma a prover informações essenciais que sejam facilmente entendidas.

6) Entrega a tempo e disseminação de relatórios: Achados significativos e relatórios de avaliação devem ser disseminados junto aos usuários intencionais da avaliação, para serem utilizados no momento adequado.

7) Impacto da avaliação: Avaliações devem ser planejadas, conduzidas e divulgadas de forma a estimular seu acompanhamento por parte dos principais interessados, aumentando assim as possibilidades de uso da avaliação.

3.2.2 Standards de Viabilidade

Os *Standards* de Viabilidade objetivam assegurar que uma avaliação será realista, prudente, diplomática e simples. São três estes *Standards*:

1) Procedimentos práticos: Os procedimentos da avaliação devem ser práticos a fim de evitar ao máximo perturbações no momento em que as informações necessárias estão sendo coletadas.

2) Viabilidade política: A avaliação deve ser planejada e conduzida tendo-se clareza antecipada dos diferentes posicionamentos dos vários grupos de interesse. Deve assim conseguir a cooperação de todos e tentar evitar, ou criar mecanismos para contra-atacar, possíveis esforços, de qualquer destes grupos, para dificultar a operacionalização, ou confundir, ou usar de forma inadequada os resultados da avaliação.

3) Custo-efetividade: A avaliação deve ser eficiente e produzir informações de valor significativo, de forma que os recursos utilizados possam ser justificados.

3.2.3 Standards de Propriedade

Estes *Standards* visam a assegurar que uma avaliação será conduzida dentro de normas legais e éticas, respeitando o bem-estar das pessoas envolvidas na avaliação e daquelas que venham a ser afetadas por seus resultados. Os oito *Standards* de Propriedade são os seguintes:

1) Orientação para o serviço: avaliações devem ser desenhadas de forma a ajudar as organizações a atender e responder de forma eficaz às necessidades de todas as pessoas e instituições servidas ou afetadas de alguma maneira por seus programas.

2) Acordos formais: acordos firmados sobre “o quê”, “como”, “quem” e “quando” fazer as atividades pertinentes à avaliação devem estar registrados num documento, de modo que as partes envolvidas sintam-se na obrigação de honrar os compromissos assumidos ou renegociá-los formalmente se for necessário.

3) Direitos dos indivíduos: avaliações devem ser planejadas e levadas a cabo considerando sempre o respeito e o zelo pelos direitos individuais das pessoas nelas envolvidas ou afetadas por seus resultados.

4) Relações humanas: os avaliadores, no momento de se relacionar com as outras pessoas envolvidas na avaliação, devem respeitar a dignidade e o valor dessas

peças, para que elas não se sintam em nenhum momento ameaçadas ou prejudicadas.

5) Avaliação completa e justa: as avaliações devem ser completas e justas no momento de examinar e registrar os principais pontos positivos e limitações do programa avaliado, possibilitando a valorização dos aspectos de sucesso do programa e a correção das falhas existentes.

6) Divulgação dos resultados: os responsáveis pela avaliação devem assegurar que todas as pessoas e instituições afetadas pelo processo avaliativo, e aquelas com direito legal, tenham acesso aos resultados da avaliação em sua totalidade, assim como a possíveis limitações relevantes do processo.

7) Conflito de interesses: qualquer conflito de interesses durante o desenvolvimento do processo avaliativo deve ser tratado de forma aberta e honesta entre as partes envolvidas, de forma a não comprometê-lo ou a seus resultados.

8) Responsabilidade fiscal: a alocação e o desembolso dos recursos destinados à avaliação devem ser realizados de forma adequada, prudente e eticamente responsável pelo avaliador, para que se possa prestar contas dos gastos de maneira transparente.

3.2.4 *Standards* de Precisão

Os *Standards* de Precisão visam a assegurar que uma avaliação irá revelar e produzir informações tecnicamente adequadas sobre os aspectos que determinam mérito e relevância do programa avaliado. Estes *Standards* são doze:

1) Documentação do programa: o programa avaliado deve ser descrito e documentado de forma clara e precisa, para que possa ser facilmente identificado.

2) Análise do contexto: o contexto no qual o programa está inserido deve ser descrito com o detalhamento necessário para que suas possíveis influências no programa possam ser identificadas.

3) Descrição de propósitos e procedimentos: os propósitos e procedimentos da avaliação devem ser descritos de maneira clara e monitorados para que sejam facilmente identificados e examinados.

4) Fontes de informação defensáveis: as fontes de informação usadas na avaliação devem ser descritas em detalhes, de forma a permitir a análise da adequação da informação coletada.

5) Informações válidas: os procedimentos para a coleta de informações devem ser escolhidos ou desenvolvidos e depois aplicados de forma a assegurar que a interpretação de resultados seja válida e útil.

6) Informações confiáveis: os procedimentos para coleta de informações devem ser escolhidos ou desenvolvidos e depois aplicados de forma a assegurar que a informação obtida seja suficientemente confiável.

7) Informação sistemática: a informação coletada, processada e escrita em relatórios deve ser sistematicamente revisada e qualquer erro encontrado deve ser corrigido.

8) Análise de informações quantitativas: informações quantitativas devem ser apropriada e sistematicamente analisadas de forma a possibilitar responder efetivamente às perguntas avaliativas formuladas.

9) Análise de informações qualitativas: informações qualitativas devem ser apropriada e sistematicamente analisadas de forma a possibilitar responder efetivamente às perguntas avaliativas formuladas.

10) Conclusões justificáveis: as conclusões de uma avaliação devem ser explicitamente justificadas para que possam ser analisadas pelos principais interessados/afetados pela avaliação e/ou pelo programa.

11) Imparcialidade de relatórios: os procedimentos para elaboração de relatórios deverão incluir métodos que previnam possíveis distorções causadas por sentimentos/opiniões/vieses pessoais, a fim de que os relatórios reflitam de forma justa os resultados da avaliação.

12) Meta-avaliação: o processo de desenvolvimento, bem como o possível impacto da avaliação para o programa, devem ser avaliados utilizando os *Standards* descritos aqui, ou outros, a fim de garantir que o processo transcorra de forma adequada e que, ao final, os principais interessados/afetados pela avaliação possam visualizar com clareza os pontos positivos e as limitações da avaliação.

3.3 PESQUISAS SOBRE A APLICABILIDADE INTERNACIONAL DOS *STANDARDS*

Alguns estudos foram realizados sobre a aplicabilidade internacional dos *Standards*.

Widmer (1995) realizou 15 meta-avaliações na Suíça, a fim de verificar a relação entre os *Standards* e a prática da avaliação no país. O autor concluiu que os *Standards* são perfeitamente aplicáveis à realidade daquele país e que o seu uso foi muito positivo para a análise das práticas avaliativas.

Jang (2000) realizou pesquisa quantitativa descritiva visando a verificar a aplicabilidade dos *Standards* na Coreia do Sul. O autor verificou uma maior aplicabilidade dos *Standards* relacionados a Viabilidade e Precisão. Foi observado um choque de valores em relação a alguns dos *Standards* de Utilidade e Propriedade (orientação para o serviço, identificação de interessados, conflitos de interesse e disseminação dos achados).

Taut (2001) realizou estudo baseado em literatura *cross-cultural* psicológica e antropológica para identificar os valores que orientam cada um dos trinta *Standards*. Segundo o autor, a origem norte-americana confere aos *Standards* características que podem limitar sua transferência para outros países. Foram utilizadas as seguintes dimensões culturais na análise: individualismo vs. coletivismo, conservadorismo vs. autonomia, hierarquia vs. igualitarismo, domínio vs. harmonia, comunicação direta vs. comunicação indireta, rejeição à incerteza, alto contexto vs. baixo contexto, e senioridade. O autor concluiu que os *Standards* de Utilidade e Propriedade têm

aplicabilidade limitada em outras culturas. Alguns dos motivos desta dificuldade decorrem do fato de que em muitos países da Ásia, como Índia e Coréia, valores como o conservadorismo e a hierarquia sobrepujam os valores norte-americanos da autonomia e do igualitarismo. O autor recomenda aos avaliadores que, trabalhando fora dos Estados Unidos, levem em consideração as dimensões culturais e sociais locais, numa análise sistemática e crítica de cada um dos *Standards*.

4 MODELAGEM DA PESQUISA

4.1 MÉTODO ADOTADO

Adotou-se como modelo orientador da pesquisa a metodologia de estudo de casos múltiplos proposta por Yin (2001). Esta escolha deu-se pela necessidade de um estudo em profundidade da adoção de determinados procedimentos sugeridos pelos *Standards*. Além da visão de que o estudo de caso seja útil para realizar pesquisas de tipo exploratório, neste estudo o método de caso foi empregado para realizar uma pesquisa descritiva. Segundo Yin, o estudo de caso tem sido muitas vezes interpretado de maneira equivocada por aqueles que propõem que diferentes estratégias de pesquisa sejam dispostas de maneira hierárquica e que, nesta perspectiva, o estudo de caso seja indicado apenas para pesquisas exploratórias. Segundo Schramm (1971) apud Yin (2001, p. 31), “a essência de um estudo de caso é tentar esclarecer uma decisão ou um conjunto de decisões, incluindo o motivo pelo qual as decisões foram tomadas, como foram implementadas e quais resultados foram alcançados”. Yin sugere uma compreensão pluralística quanto às possíveis estratégias para o emprego do estudo de caso e propõe três modalidades possíveis: estudos de casos exploratórios, descritivos e explanatórios.

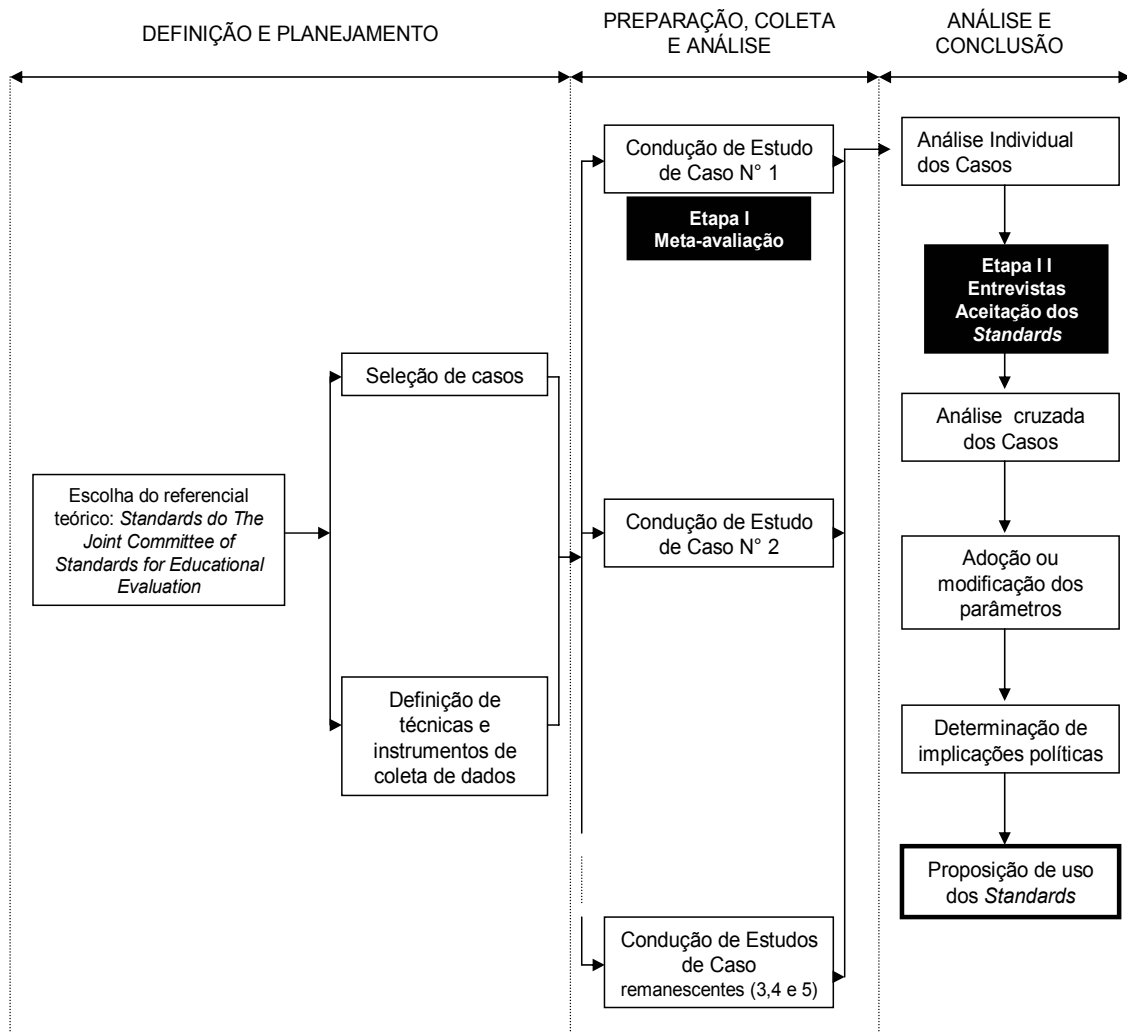
Gil (2002) aponta três objeções para o uso do estudo de caso único nas ciências sociais. A primeira delas diz respeito à ausência de rigor metodológico quando comparado com modelos experimentais ou com levantamentos; esta fragilidade se traduz em viés que pode comprometer o resultado das pesquisas. A segunda objeção refere-se à dificuldade de generalização, já que é impossível propor generalizações a partir do estudo de um ou de poucos casos. A terceira objeção diz respeito ao longo período de tempo demandado para este tipo de pesquisa que, segundo o autor, gera resultados pouco consistentes em função das mudanças rápidas no meio social.

Na presente pesquisa, pretendeu-se contornar as objeções com a adoção dos seguintes procedimentos: 1. Quanto à possibilidade de viés, pretendeu-se minimizar sua ocorrência fazendo-se uso de instrumento quantitativo combinado com roteiros qualitativos para coleta de informações; 2. Quanto à dificuldade de generalização, utilizou-se o estudo de um conjunto restrito de casos em função da dificuldade de identificar um número significativo de avaliações no setor pesquisado. Como já foi citado (IPEA, 2000 e GIFE, 2000), raras são as organizações que realizam avaliações sistemáticas; e, 3. Quanto ao fator tempo, procurou-se selecionar casos cuja natureza e acesso permitissem que os procedimentos de coleta, análise e interpretação fossem realizados de forma relativamente ágil.

O método de casos múltiplos proposto por Yin foi adaptado em alguns aspectos a fim de servir como base lógica para o estudo. O principal aspecto modificado diz respeito à etapa inicial, que prevê o desenvolvimento de uma teoria ou padrão a ser pesquisado nos casos (figura 3). Foram escolhidos para ocupar esse papel os *Standards* de Avaliação de Programas, que não constituem uma teoria, mas o paradigma vigente de avaliação de programas nos Estados Unidos. Segundo Campbell (1975) apud Yin (2001), a adequação a determinado padrão pode ser entendida como o

meio pelo qual várias partes da mesma informação do mesmo caso podem ser relacionadas à mesma proposição teórica; enquanto paradigma, os *Standards* sugerem padrões passíveis de comparação que se originaram de várias teorias e práticas em uso.

Figura 3 – Método de estudo de casos múltiplos adaptado



Fonte: COSMOS Corporation apud Yin (2001), pg. 73

As justificativas quanto à definição amostral, dados e procedimentos de análise nas etapas da pesquisa são descritas a seguir. O trabalho de campo compreendeu dois momentos de coleta de dados (Etapas I e II). Portanto, além de uma coleta proposta pelo método original, foi realizada uma segunda, que precedeu a análise cruzada dos dados.

4.2 SELEÇÃO DOS CASOS

Sendo esta uma pesquisa exploratória que não busca a incidência ou a predominância de fenômenos, não é recomendável o uso de procedimentos probabilísticos de definição amostral. Foram selecionadas cinco organizações que fazem investimento social privado e que se utilizam da avaliação de programas e projetos como componentes de seu sistema de gestão. Definiu-se como “casos”, para o contexto desta pesquisa, processos de avaliação externa de programas operados ou apoiados pelas organizações que fizeram parte da amostra. Foram selecionados casos que atendessem aos seguintes critérios: 1. A avaliação ocorreu nos últimos três anos; 2. O gerente que contratou a avaliação vivenciou o processo desde a contratação até a sua finalização; e, 3. Existe documentação necessária para coleta de dados (contrato da avaliação e relatório da avaliação). Além desses critérios, foi sugerido à instituição que contratou mais de uma avaliação que indicasse para ser “caso” da pesquisa aquela que considerava ser a melhor delas realizada até aquele momento.

As organizações gestoras dos casos de avaliação selecionados foram:

- Fundação ABRINQ
- Fundação Telefônica
- Fundação Vale do Rio Doce
- Instituto Ayrton Senna
- Raytheon Corporation

4.3 DETALHAMENTO DA PRIMEIRA ETAPA DA PESQUISA

A pesquisa foi realizada em duas etapas visando, a permitir levantar, em um primeiro momento, uma análise dos procedimentos adotados desde a elaboração da proposta de avaliação pelos avaliadores até a entrega de relatórios. Somente a partir dessa análise é que foi possível estruturar um quadro que configurasse o uso dos *Standards* enquanto critérios de julgamento das práticas adotadas para que se pudesse, então, passar a buscar respostas para a questão central da pesquisa.

Para a primeira etapa, que buscou respostas para a pergunta “Em que medida as avaliações externas de programas e projetos de organizações do terceiro setor obedecem aos *Standards*?”, foram realizadas análises de documentos e aplicados instrumentos de *Checklist* para estruturação e análise dos casos individualmente.

Para estudar as práticas avaliativas dos casos com os *Standards* fez-se uso da meta-avaliação. Esta técnica foi proposta por Scriven (1969) apud Stufflebeam (1974) e refere-se à avaliação de avaliações. A técnica foi concebida para avaliar eventuais problemas como vieses, questões administrativas e erros técnicos. Segundo Worthen, Sanders & Fitzpatrick (1997, p. 440), durante a década de 60, avaliadores começaram a discutir a necessidade de meta-avaliações formais para garantir as distinções entre as avaliações bem feitas das mal feitas. A partir desse período, vários avaliadores começaram a publicar suas propostas de critérios para meta-avaliação (Stake, 1969; Stufflebeam et al., 1971; Stufflebeam, 1974; Scriven, 1974; Rossi, 1982). Os esforços para elaboração de critérios para avaliar avaliações convergiram para a elaboração dos *Standards* de Avaliação de Programas.

Stufflebeam (2000) desenvolveu um instrumento denominado *Program Evaluation - Meta Evaluation Checklist*, com base nos *Standards* de Avaliação de Programas. O instrumento é indicado para meta-avaliações e foi adaptado para o presente estudo, visando à aplicação para a avaliação dos casos selecionados. É composto pelos 30 *Standards* já apresentados. Para cada um deles são propostos seis itens (*checkpoints*) a serem verificados (ver tabela 3), o que significa 180 itens a serem verificados em cada caso. A adaptação feita no instrumento foi a inclusão de uma escala intermediária quanto ao cumprimento dos *checkpoints*. O instrumento previa apenas duas escalas (cumpriu e não cumpriu). Na realização do pré-teste, percebeu-se que a escala proposta era muito rígida, incluindo-se na versão utilizada neste estudo uma opção intermediária (cumpriu parcialmente).

Tabela 3 - Exemplo de um *Standard* e seus *checkpoints* específicos:*

U1 Identificação dos interessados		
	Foram definidos com o cliente os interessados mais importantes	
	O avaliador engajou líderes para identificar outros interessados	
	Os interessados ajudaram a identificar outros interessados	
	Foram consultados interessados para a identificação de suas necessidades de informação	
	Foram propostas estratégias para envolvimento dos interessados durante a avaliação no contrato	
	O avaliador manteve o processo de avaliação aberto para a inclusão de novos interessados	
1- cumpriu o item	0,5 – cumpriu parcialmente	0- não cumpriu o item
Total =		

* O instrumento completo encontra-se no Anexo A.

Fonte: Autor

O instrumento foi aplicado em duas etapas. Na primeira foi feita análise de documentos (proposta de consultoria, contrato, relatórios parciais e relatório final) para a aplicação de *Checklist*. Para alguns *Standards* não havia informação suficiente

nos documentos analisados para se proceder à verificação. Para se obter essas informações, foi solicitado aos gerentes das organizações envolvidas na pesquisa e avaliadores o preenchimento de partes do instrumento sobre informações não encontradas nos documentos analisados.

4.3.1 Análise e interpretação dos dados da primeira etapa

Apesar do instrumento de *Checklist* possibilitar um cômputo estatístico do conjunto de casos quanto à frequência e à prevalência de situações de não cumprimento, cumprimento parcial ou cumprimento dos *checkpoints* pelo grupo de casos estudados, as análises numéricas realizadas procuram obedecer a critérios qualitativos quanto à replicação de situações de não cumprimento ou de cumprimento parcial dos *Standards*.

Para Yin (2001), a aplicação da lógica estatística não se aplica nos estudos de caso por três motivos: 1) Os estudos de caso não são indicados para avaliar a incidência de fenômenos apenas; 2) Um estudo de caso deve tratar do fenômeno e do contexto onde ele se insere; e, 3) Cada estudo de caso consiste em um estudo completo, no qual buscam-se provas convergentes a respostas de fatos que orientam conclusões de casos individuais. Em um conjunto de casos, as conclusões individuais necessitam de replicação nos outros casos para serem consideradas consistentes.

A análise matemática dos *Checklist* objetivou buscar em cada caso estudado, situações de convergência com relação ao cumprimento dos *checkpoints* e de divergência quanto a situações de não cumprimento ou cumprimento parcial dos *Standards*. A análise cruzada dos casos possibilitou a identificação de replicação das três possíveis situações.

Os instrumentos foram analisados individualmente, utilizando-se o critério de soma dos *checkpoints* cumpridos ou cumpridos parcialmente, possibilitando uma pontuação de 0 (não cumprimento dos *checkpoints*) a 6 (cumprimento total dos

checkpoints) em cada um dos *Standards*. Após a análise individual, foi realizada análise do conjunto das meta-avaliações, utilizando-se as médias de pontuação dos casos nos *Standards*. As médias encontradas foram convertidas em percentuais, para orientar decisões analíticas para a segunda etapa da pesquisa.

O produto gerado pela análise desta primeira etapa foi uma relação de 17 *Standards* definidos como críticos por não terem sido cumpridos totalmente por pelo menos três dos cinco casos.

Além da análise em percentuais, efetuou-se uma análise de desempenho do conjunto de casos segundo a escala proposta por Stufflebeam (2000). A escala varia de Pobre, Ruim, Bom, Muito Bom e Excelente. O que determina o desempenho nesta proposta é o maior número de *checkpoints* cumpridos.

4.4 DETALHAMENTO DA SEGUNDA ETAPA DA PESQUISA

Na segunda etapa da pesquisa foram realizadas entrevistas individuais e aplicação de questionários a cinco gerentes e cinco avaliadores visando a aprofundar o estudo dos casos individualmente e estabelecer uma discussão sobre a adequação dos critérios das bases de julgamento utilizadas (os *Standards*). Foram formulados roteiros de entrevista estruturados a partir do produto gerado na primeira etapa da pesquisa.

Os roteiros para gerentes e avaliadores diferiram em alguns aspectos. Parte do roteiro dos gerentes cumpriu o papel de assegurar a validade interna da pesquisa. Foram solicitadas informações sobre fatos e procedimentos referentes aos itens abordados pelo instrumento de *Checklist*, com o intuito de confirmar a análise de documentos. Segundo Patton (2002), este procedimento é chamado de triangulação de fontes de informação. Para os avaliadores, respeitando o *status* de especialistas, solicitou-se comentários sobre o não cumprimento dos *checkpoints* dos *Standards* críticos pelos casos. Na parte comum do roteiro, os entrevistados foram convidados a opinar sobre sua concordância ou discordância quanto à aplicabilidade dos *Standards*. Com isto, o processo de entrevista assegurou uma discussão prévia sobre o uso dos parâmetros propostos com base nas práticas dos entrevistados.

4.4.1 Análise e interpretação dos dados da segunda etapa

Em um primeiro momento, foi realizada a análise de percentuais no que se refere à concordância quanto aos *Standards* propostos. A partir disso, selecionou-se o grupo que obteve menor concordância para análise dos dados qualitativos levantados nas entrevistas. Após a organização das transcrições, utilizou-se a técnica de análise de conteúdo, com o objetivo de identificar temas e tendências sobre as questões discutidas nas entrevistas. O método de análise utilizado foi o indutivo. Segundo Patton (2002), a análise indutiva envolve a descoberta de termos, temas e categorias em um conjunto de dados. O resultado da análise serviu de base para a elaboração de conclusões e recomendações.

5 ANÁLISE E DISCUSSÃO

5.1 ANÁLISE DESCRITIVA

5.1.1 Características das instituições

Na tabela 4 são apresentadas as características gerais das organizações participantes da pesquisa.

Tabela 4 – Resumo das características das organizações participantes da pesquisa

Nome	Áreas de atuação	Tempo de atuação na área social no Brasil
Fundação ABRINQ para os Direitos das Crianças	Cultura Educação Família e comunidade Saúde e nutrição Trabalho infantil	13 anos
Fundação Telefônica	Educação Saúde Desenvolvimento social	4 anos
Fundação Vale do Rio Doce	Ensino fundamental (da 1ª à 8ª série)	35 anos
Instituto Ayrton Senna	Sobrevivência - nutrição e saúde Desenvolvimento - educação, esporte, arte e profissionalização Integridade física, moral e psíquica	7 anos
<i>Rhaytheon Corporation</i> ³¹	Educação Meio ambiente	5 anos

Fonte: Autor

³¹ O investimento social da Rhaythen Corporation vem ocorrendo no âmbito da implantação do SIVAN – Sistema de Vigilância da Amazônia. A Empresa realiza essa ação por meio de seu Departamento de Relações com a Comunidade.

5.1.2 Características dos programas e das avaliações realizadas

Caso 1 - Programa Acelera Goiás – PAG (Instituto Ayrton Senna)

O programa

O PAG foi implementado em 1999 pelo Instituto Ayrton Senna, com apoio financeiro da empresa TCO Celular, em parceria com a Secretaria Estadual de Educação de Goiás, em Escolas Públicas de 193 Municípios, com o propósito de contribuir para a diminuição da defasagem idade/série, por meio do apoio a 9324 alunos multirrepetentes na superação de suas experiências escolares negativas. A principal estratégia utilizada pelo Programa é o treinamento de professores em duas metodologias, uma voltada para corrigir a distorção idade/série e outra destinada à alfabetização. O Programa preconiza a relação personalizada do professor com o aluno e o envolvimento da família no processo de aceleração de aprendizagem, para que as crianças beneficiadas se recuperem.³²

³² Fonte: folder institucional.

Características da avaliação realizada

A avaliação realizada foi externa e teve a intenção de verificar o impacto do Programa, buscando respostas às seguintes perguntas:

1. Qual o impacto do PAG sobre o desempenho dos alunos no ano de 2001?
2. Os alunos do PAG apresentam desempenho semelhante ao dos alunos da 4ª série do ensino fundamental dos cursos regulares da rede oficial, tendo como base de comparação a escala SAEB³³/99?
3. O desempenho dos alunos do PAG de 2001 foi semelhante ao desempenho dos alunos de 1999 e 2000?
4. A avaliação interna feita pelos professores do PAG correspondeu ao seu desempenho na avaliação externa?
5. Quais as características dos alunos que se submeteram ao Programa em 2001 e como elas interferem no processo de aprendizagem?

O estudo avaliativo realizado foi quantitativo, tendo sido aplicado no início e no final do período de vigência do PAG uma prova em uma amostra de 1396 educandos entre 9 e 19 anos de idade. Os resultados das provas são comparados entre si e com outros parâmetros de avaliação do desempenho dos alunos.³⁴

³³ Sistema Nacional de Avaliação da Educação Básica

³⁴ Fonte: Relatório final da avaliação.

Caso 2 – Programa Crer Para Ver – PCPV (Fundação ABRINQ para os Direitos das Crianças)

O Programa

O PCPV foi criado em 1995 a partir de uma parceria entre a Fundação Abrinq e a empresa Natura Cosméticos, com o propósito de apoiar projetos de Organizações Não Governamentais voltados à melhoria da qualidade da escola pública. A Fundação Abrinq desempenha o papel de operadora do Programa e a Natura Cosméticos é a financiadora. Um Comitê Diretor formado por representantes das duas organizações parceiras define as políticas de investimento. A aprovação dos projetos apoiados pelo PCPV é realizada sob orientação de um Comitê Técnico composto por especialistas em educação, que faz a seleção dos projetos. Entre 1996 e 2001, 127 projetos foram apoiados em todo o Brasil. Os principais temas propostos pelos projetos são: qualificação de professores; valorização da diversidade cultural; formulação e implantação de projetos político-pedagógicos; participação da comunidade na escola; protagonismo de crianças e adolescentes e construção de práticas educativas inovadoras.³⁵

Características da Avaliação realizada

A avaliação foi realizada no ano de 2000, por avaliadores externos, com propósito de verificar a efetividade social do PCPV.

³⁵ Fonte: Caderno Crer Para Ver (2002). Vol. 6 Fundação Abrinq pg. 2.

Objetivos específicos da avaliação:

1. Discutir e definir um referencial sobre a “qualidade da educação”;
2. Mapear projetos, de acordo com aspectos básicos do referencial;
3. Analisar os resultados do mapeamento, segundo critérios do Programa;
4. Analisar os relatórios do Programa, verificando se os mesmos apresentam indicações que possam dizer a favor da efetividade dos mesmos;
5. Construir um referencial para a avaliação do Programa;
6. Analisar relatórios de acompanhamento dos projetos, segundo os critérios selecionados;
7. Realizar “visitas de campo” visando à análise de uma amostra de 12 projetos apoiados pelo Crer para Ver, em diferentes regiões do País;
8. Elaborar sínteses avaliativas resultantes das análises sobre as diversas evidências.

Para levar a cabo os objetivos propostos, optou-se por uma avaliação qualitativa, por essa abordagem mostrar-se mais sensível às diferenças entre os projetos, aos acontecimentos imprevistos e às mudanças de percurso. Procurou-se avaliar os processos e os produtos de cada um dos projetos, buscando informações junto aos seguintes atores envolvidos no Programa: Comitê Técnico, Gerentes do Programa e Coordenadores e Educadores dos projetos apoiados.³⁶

³⁶ Fonte: Contrato de Prestação de serviços assinado entre a Fundação Abrinq e os responsáveis pela avaliação (2000).

Caso 3 - Rede SACI – Solidariedade, Apoio, Comunicação e Informação – RS

(Fundação Telefônica)

O Programa

A Rede Saci é um projeto que vem sendo implementado pela Coordenadoria Executiva de Cooperação Universitária e de Atividades Especiais da Universidade de São Paulo, em parceria com a Rede Nacional de Ensino e Pesquisa, o Núcleo de Computação Eletrônica da Universidade Federal do Rio de Janeiro e o Instituto Amankay de estudos e pesquisas. Além desses parceiros, que respondem pela operacionalização do projeto, a Fundação Telefônica e Vitae – Apoio à Cultura, Educação e Promoção Social patrocinam a iniciativa. A RS tem como missão: “incentivar o protagonismo e inclusão da pessoa portadora de deficiência, através do estímulo à criação de condições de acessibilidade, do compartilhamento de informações sobre deficiência e da disponibilização de ferramentas em meio digital”. A RS oferece cursos de alfabetização digital, acesso gratuito à Internet, e realiza testes de programas e equipamentos para pessoas portadoras de deficiência (PPDs) em 5 Centros de Informação e Convivência (CIC); oferece vários produtos e serviços em um *Website* (*software* adaptados, notícias, base de dados, oportunidades de trabalho, fóruns, etc.) para quase 2700 usuários (PPDs, familiares, profissionais da área, professores do ensino médio e fundamental e outros).³⁷

³⁷ Proposta de avaliação externa da Rede SACI – Solidariedade, Apoio, Comunicação e Informação. Dezembro de 2001.

Características da avaliação realizada

A avaliação da RS foi realizada em 2001 por avaliadores externos com o objetivo de verificar os principais pontos fortes e as fragilidades nas atividades desenvolvidas, e de identificar tecnologias passíveis de serem transferidas para outras iniciativas de inclusão de PPDs na sociedade. As seguintes perguntas foram formuladas para orientar o estudo:

1. Em que medida os cursos oferecidos pela Rede estão sendo capazes de desenvolver as habilidades necessárias para a inclusão de sua clientela na sociedade?
2. Qual o nível de satisfação da clientela com os cursos oferecidos e qual a possibilidade de replicar os cursos oferecidos pelo CIC?
3. Qual o nível de utilização, efetividade e satisfação das Pessoas Portadoras de Deficiência (PPDs) que acessam o *site* da Rede SACI?

No estudo avaliativo foram utilizados métodos quantitativos e qualitativos. Foi desenvolvido um *design* que previu diferentes grupos amostrais pertencentes aos distintos públicos. As formas de coleta de informações adotadas foram: questionários via *Web*, questionários via e-mail, grupos focais, observação participante, entrevistas presenciais, entrevistas por telefone e coleta de documentos impressos e eletrônicos.

Caso 4 – Programa Escola que Vale – PEV (Fundação Vale do Rio Doce)

O Programa

O PEV foi criado em 1999, a partir de uma parceria entre a Fundação Vale do Rio Doce - FVRD e a organização não governamental Centro de Estudos e Documentação para Ação Comunitária – CEDAC. A FVRD é responsável pela atuação social da Companhia Vale do Rio Doce – CVRD, a financiadora e gestora do PEV. O CEDAC, que possui larga experiência na concepção e implantação de programas sociais e educacionais, é responsável pela coordenação geral do PEV. As duas instituições selecionaram seis municípios onde a CVRD atua (Marabá e Paraupébas, PA; São Luís e Açailândia, MA; João Leiva, ES; e, Catas Altas, MG) e realizaram um diagnóstico da situação das escolas públicas envolvendo diferentes atores da comunidade escolar. A partir do diagnóstico definiu-se o seguinte propósito para o PEV: “provocar uma transformação substancial da prática do professor e, conseqüentemente, da escola, permitindo a professores e alunos estabelecerem uma nova relação com o conhecimento... trazer aos professores e alunos a possibilidade de eleger conteúdos em sentido real e sugerir trabalhos que tenham significado para quem os realiza e para a comunidade de dentro e de fora da escola”³⁸. A principal estratégia utilizada para levar a cabo o propósito do Programa foi a capacitação dos professores de 25 escolas públicas nos seis municípios na implantação de projetos educativos. O CEDAC ofereceu 13 alternativas temáticas de projetos para a escolha pelos professores e posterior implantação nas escolas.

³⁸ Fonte: Documento de logística. Janeiro de 2000. p. 31 apud Relatório final da avaliação. p. 6.

Características da avaliação realizada

A avaliação realizada, denominada por seus responsáveis como “pesquisa avaliativa”, foi externa, processual e teve como principal propósito “aprender os processos de implementação do Programa, assim como os fatores e relações facilitadoras ou complicadoras na obtenção dos resultados pretendidos, considerando que tais resultados não se esgotam na capacitação do professor e nem em 6 meses de intervenção”³⁹. Os seguintes eixos estruturantes compuseram o *design* do estudo realizado:

- **Os alunos:** ganhos no desempenho escolar, em sociabilidade e participação na vida pública/comunitária;
- **O professor:** ganhos de competência e de prática democrática; impactos na vida pessoal/profissional e na vida pública/comunitária;
- **A escola:** qualidade do vínculo escola/comunidade e ganhos de competência e inserção na comunidade;
- **As famílias:** envolvimento e impactos na ampliação de seu universo informacional/cultural e na participação pró-ativa no âmbito da escola;
- **A comunidade:** grau de envolvimento e impactos no fortalecimento de vínculos e de identidade de seus habitantes.
- **A rede local de parcerias:** seu desempenho referido aos processos de mobilização e irradiação na melhoria da qualidade do ensino municipal.

³⁹ Fonte: Pesquisa Avaliativa Programa “Escola que Vale”, São Paulo, 2000

A pesquisa avaliativa teve caráter quantitativo e qualitativo. Foram coletadas, em uma primeira etapa, informações relativas ao contexto socioeducacional dos seis municípios, através do levantamento de alguns indicadores relacionados aos interesses de atuação do PEV. Na segunda etapa do estudo, avaliou-se a percepção dos professores quanto aos diferentes Projetos de Trabalho desenvolvidos, em relação aos eixos estruturantes propostos. A terceira e última etapa envolveu a aplicação de uma “Ficha de Atitudes e Habilidades” dos alunos, preenchidas pelos professores.⁴⁰

⁴⁰ Fonte: Pesquisa Avaliativa Programa “Escola que Vale”. (Relatório Parcial), 2000.

Caso 5 – Programa de Educação na Amazônia – PEA

(Raytheon Corporation)

O Programa

O PEA foi criado em 1998 pelo Departamento de Relações com a Comunidade da Raytheon Corporation, vinculado à implantação do Projeto SIVAN – Sistema de Vigilância da Amazônia. Um diagnóstico realizado por uma consultoria especializada sugeriu as seguintes orientações para o desenho do PEA: 1. centrar as ações do Programa em formação continuada de professores do ensino fundamental de escolas públicas; 2. incluir as Universidades Federais nos projetos; 3. contratar um gerente para monitorar a implementação dos projetos; 4. localizar os projetos nos três grandes centros urbanos da Amazônia, ou seja, as cidades de Belém, Manaus e Porto Velho, que são os centros regionais do SIVAM. Em 1999 foram constituídas três equipes das Universidades Federais das cidades envolvidas, que desenvolveram projetos visando a atender demandas específicas dos sistemas educacionais públicos dos três municípios, atendendo a quatro eixos comuns: 1) prioridade para as primeiras quatro séries do Ensino Fundamental das escolas públicas; 2) capacitação dos professores em serviço; 3) ações complementares que proporcionem um envolvimento das comunidades em torno das escolas participantes; e, 4) produção de resultados mensuráveis. Os projetos envolveram a participação de 492 profissionais, entre coordenadores, docentes das universidades, diretores e professores de escolas públicas.

Características da avaliação realizada

A avaliação foi externa, de processo e teve como propósito: “detectar pontos fortes e frágeis, bem como oportunidades e riscos de cada Projeto, a exemplo do formato *SWOT* – *Strengths, Weaknesses, Opportunities, Threats...* para consolidar entendimentos, apoiar necessárias atuações e ampliar comprometimento e aperfeiçoamento de indivíduos, grupos, projetos, programas, instituições e sistemas, enquanto permite a formulação de juízos e recomendações, que geram ações, políticas e conhecimentos.”⁴¹ Foram formuladas três perguntas para orientar o estudo avaliativo:

1. Até que ponto o Programa de Educação na Amazônia, através de seus Projetos, evidencia resultados significativos e justificáveis?
2. O Programa de Educação na Amazônia, através de seus Projetos, vem desenvolvendo um processo efetivo para o alcance de seus propósitos?
3. Até que ponto o Programa de Educação na Amazônia, através de seus Projetos, evidencia capacidade de auto-sustentação e continuidade?

⁴¹ Fonte: Relatório final da avaliação do Programa Educação na Amazônia, 2002,p. 6.

A abordagem metodológica adotada foi baseada no modelo construtivista e responsivo proposto por Guba & Lincoln. Houve intenso envolvimento dos interessados desde a definição das perguntas avaliativas e indicadores, até a fase de discussão dos resultados preliminares da avaliação. Além do desenvolvimento de estreito relacionamento com os interessados, foram coletadas informações quantitativas e qualitativas com os diferentes atores do PEA, através de grupos focais, entrevistas individuais, *Checklists*, conversas informais, observação participante e consulta a documentos.⁴²

⁴² Fonte: Relatório final da avaliação do Programa Educação na Amazônia, 2002,p. 6.

5.1.3 Perfil dos Entrevistados

A. Perfil dos gerentes

Foram entrevistados cinco gerentes, responsáveis pelos programas que passaram pela avaliação, acompanhando de perto todas as etapas do processo. Quatro gerentes entrevistados estão entre 35 e 50 anos de idade, um outro está acima de 50 anos de idade. Os cinco possuem terceiro grau, sendo que um deles possui o grau de mestre e outro o de doutor. Todos têm formação na área de ciências humanas. Os cinco gerentes estão há mais de três anos gerenciando os programas e acompanharam todo o processo de avaliação realizado. Três gerentes estão há pelo menos três anos na instituição, e dois há mais de seis anos. Todos possuem experiência prévia com trabalho na área social. As características de formação, experiência profissional e tempo à frente dos programas denotam um alto grau de profissionalização e de conhecimento sobre estes programas e sobre as avaliações alvo desta pesquisa.

B. Perfil dos Avaliadores

Os cinco avaliadores entrevistados, responsáveis pela coordenação das avaliações realizadas, estão acima dos 50 anos de idade. Quatro são doutores e um é livre-docente. Quatro possuem graduação na área de ciências humanas e um na área de ciências biológicas. Todos exercem atividades de docência e pesquisa. Quatro atuam há mais de 20 anos nessas atividades e um deles entre 10 e 15 anos. A experiência, especificamente na área de avaliação, também é extensa. Três atuam nesta área há mais de 20 anos e dois entre 10 e 15 anos. Os entrevistados possuem portanto, extensa e respeitável experiência em docência, pesquisa e avaliação. Nas entrevistas foi possível constatar que as atividades de avaliação de programas e políticas sociais governamentais ocuparam preponderantemente a pauta de trabalho desses pesquisadores. Conforme sua experiência, a demanda por avaliação de projetos e programas do terceiro setor vem crescendo nos últimos cinco anos.

5.2 ANÁLISE DE RESULTADOS - PRIMEIRA ETAPA DA PESQUISA - META-AVALIAÇÃO

A seguir são apresentadas as análises de desempenho do conjunto de casos em relação aos 180 *checkpoints* verificados nos casos individualmente. Os resultados desta meta-avaliação são demonstrados em termos percentuais de cumprimento dos itens e em relação aos critérios originalmente propostos por Stufflebeam (2000) no instrumento de *Checklist*. As análises permitem responder à seguinte questão:

Questão 1 - Em que medida as avaliações externas de programas e projetos de investimento social privado obedecem aos *Standards*?

Na tabela 5 é apresentado o desempenho dos cinco casos estudados em relação aos 30 *Standards* em termos percentuais. Apesar da representação numérica, não se pretendeu com os números realizar uma análise de frequência de fenômenos, e sim, a partir das informações quantitativas, identificar convergências de atendimento aos padrões propostos pelos *Standards*.

Além da representação em percentuais, na figura 4 e na tabela 6 são apresentados os desempenhos médio e individual dos casos, seguindo a proposta original do instrumento utilizado. Na tabela 7 está representada a identificação de “*Standards* Críticos”, definidos a partir de uma das análises do *Checklist*, observando-se a incidência no grupo de situações de não cumprimento e/ou cumprimento parcial dos *checkpoints* por três ou mais casos.

5.2.1 Desempenho geral dos casos em relação aos *Standards*

O conjunto dos *Standards* de Utilidade - U foi cumprido em menor medida (76%), seguido do de Precisão - PC (77%), Viabilidade - V (81%) e Propriedade - PR (81%). Apesar dos elevados percentuais, ao aplicar os critérios de análise de Stufflebeam, utilizando a média de pontuação do conjunto de casos em cada um dos *Standards*, observou-se um desempenho “Bom” em cada um dos grupos (tabela 6). Esse desempenho esteve muito próximo de um patamar maior “Muito Bom” nos *Standards* de Utilidade, Viabilidade e Precisão.

Vale ressaltar que a análise realizada nesta etapa destina-se a fornecer elementos para a segunda etapa da pesquisa. Conforme acordado com as instituições envolvidas com os casos estudados, não houve nesta etapa o intuito de emissão de juízo de valor sobre as avaliações realizadas. O principal objetivo foi estabelecer um quadro de situação dos processos avaliativos alvo da pesquisa em relação aos *Standards*.

A. Desempenho dos casos em relação aos *Standards* de Utilidade (U)

Em três dos sete *Standards* de Utilidade (*U1 - Identificação de interessados*: 50%; *U6 - Entrega a tempo e disseminação de relatórios*: 55%; e, *U7 - Impacto da avaliação*: 67%), os casos apresentaram desempenho visivelmente inferior aos outros quatro. O baixo desempenho no *U1 - Identificação de interessados*, refere-se a uma característica de quatro dos cinco casos na identificação e atendimento de

necessidades, apenas de interessados ligados à organização financiadora da avaliação e/ou do grupo gestor do Programa. A diretriz proposta pelo U1 é “Pessoas envolvidas ou afetadas pela avaliação devem ser identificadas, de forma que suas necessidades possam ser atendidas”, não se restringindo, portanto, aos interesses apenas dos que contratam a avaliação. O U1 estende a outros usuários do Programa, que são em alguma medida afetados pela avaliação, a possibilidade de ter suas questões respondidas por ela. Em apenas um dos casos observou-se o engajamento de outros interessados além dos patrocinadores e gestores do Programa.

Com relação ao *U6 - Entrega a tempo e disseminação de relatórios*, dois aspectos contribuíram para o baixo desempenho. Apenas um dos casos cumpriu os prazos previstos para a entrega dos relatórios. Nenhum dos casos cumpriu totalmente um *checkpoint* que diz respeito ao desenvolvimento de *releases* para divulgação na mídia. O primeiro aspecto é o mais crítico com relação ao cumprimento do U6, já que o cumprimento dos prazos na produção de relatórios está relacionado ao fornecimento de informações para a tomada de decisões, como observa-se no enunciado do U6: “relatórios parciais com informações relevantes devem ser produzidos e disseminados durante o processo avaliativo para que decisões importantes possam ser tomadas em tempo oportuno”. A morosidade no preparo de relatórios e na disseminação das informações compromete um atributo essencial de uma avaliação que é a geração de informações úteis à tomada de decisões.

O baixo desempenho no *U7 Impacto da avaliação*, deveu-se ao não cumprimento total dos *checkpoints* pelo conjunto de casos em três itens. O primeiro item está relacionado ao fornecimento de *follow-up*, suporte e apoio para a interpretação dos achados da avaliação. Dois casos cumpriram totalmente esse item, um não cumpriu e dois outros cumpriram parcialmente. O segundo item está

relacionado à clareza com a qual o avaliador comunicou as possibilidades de incorporação dos achados no trabalho dos beneficiários da avaliação. Um dos casos cumpriu plenamente esse item, um não cumpriu e três cumpriram parcialmente. O terceiro *checkpoint* diz respeito a reuniões de *feedback* para discussões mais aprofundadas, para ir além da interpretação e discutir a implementação das recomendações geradas pela avaliação.

B. Desempenho dos casos em relação aos *Standards* de Viabilidade (V)

O *Standard* de menor desempenho foi o V3 – *Custo efetividade* (73%). Um dos casos cumpriu todos os *checkpoints*. Nos outros quatro casos, quatro foram os *checkpoints* nos quais observou-se menor desempenho. O primeiro deles está relacionado à produção, pela avaliação, de informações novas, úteis à tomada de decisões. Um caso cumpriu totalmente o item e três parcialmente. Nos três outros *checkpoints*, repetiu-se o cumprimento observado no anterior (um caso cumpriu totalmente e três parcialmente). Os três *checkpoints* referem-se a: 1) Fomento pelo avaliador a melhorias do Programa; 2) Produção de informações sobre o *accountability* (uso responsável de recursos e prestação de contas às autoridades superiores ligadas ao programa. Exemplo: Financiadores e Conselho Diretor); e, 3) Geração de novos *insights* sobre o programa.

O V3, cuja diretriz é “a avaliação deve ser eficiente e produzir informações relevantes de modo a justificar os custos”, diz respeito à qualidade das informações geradas pelo processo avaliativo. Enquanto os dois *checkpoints* dos *Standards* (U6 - *Entrega a tempo e disseminação de relatórios* e U7 - *impacto da avaliação*), dizem respeito a procedimentos para assegurar o uso da informação, a dimensão do

custo/efetividade sugere a dimensão da relação esforço (tempo e recursos) com a dimensão qualidade do produto (informações novas, úteis e geradoras de aprendizagens). A fragilidade nesses quesitos pode estar relacionada ao entendimento que se estabelece entre avaliador e cliente quando se define o objeto, o foco e os usos que se pretendem fazer das informações geradas pela avaliação.

C. Desempenho dos casos em relação aos *Standards* de Propriedade (PR)

O PR2 – *Termos contratuais previstos*, foi o de menor desempenho pelo conjunto de casos (67%). Dois foram os *checkpoints* nos quais os casos apresentaram um maior número de não cumprimento ou cumprimento parcial. *Definição do formato de relatórios* foi cumprido totalmente por apenas um dos casos, parcialmente por um outro e três deles não cumpriram esse item. *Definição de estratégias de disseminação* foi cumprido parcialmente por um caso e não cumprido pelos outros quatro casos. Essas constatações revelam pouca preocupação por parte dos avaliadores e também a não observação por parte dos gerentes da necessidade de se definir estratégias de comunicação *a priori*.

O desempenho dos casos com relação ao PR5 – *Levantamento completo e justo* (73%) foi afetado pelo não cumprimento pela maioria deles dos *checkpoints* relacionados ao não reconhecimento no relatório final, por parte dos avaliadores, das limitações na análise de alguns aspectos e do processo avaliativo. Em dois dos cinco relatórios foram expressas informações sobre as limitações dos estudos avaliativos efetuados. Com relação às limitações nas análises, apenas um caso cumpriu parcialmente o item. Na análise dos *Standards* de Precisão a discussão desses itens será retomada.

Nos três outros *Standards* de menor percentual (PR6 - 83%, PR7 - 78% e PR8 - 68%), apenas um *checkpoint* em cada um deles influenciou um percentual mais baixo no cumprimento dos *Standards*. Com relação ao PR6 – *Divulgação dos achados*, observou-se que os cinco casos cumpriram parcialmente o item relacionado ao empenho do avaliador para que os relatórios chegassem a diferentes audiências. O compartilhamento restrito de informações está estreitamente relacionado ao envolvimento de poucos interessados desde o início do processo avaliativo. O item de menor cumprimento, o PR7 – *Conflitos de interesse*, está também relacionado ao encaminhamento de relatórios, mas, para uma audiência específica, ou seja, o diretor presidente da organização. Em dois dos cinco casos houve por iniciativa do avaliador o envio de relatório diretamente para o diretor presidente da organização contratante. O *checkpoint* relacionado à inclusão de um informe sobre o uso dos recursos destinados à avaliação pelos avaliadores foi o item que mais influenciou o baixo percentual relativo ao cumprimento do PR8 – *Responsabilidade fiscal*. Em nenhum dos cinco casos observou-se a inclusão de informes dessa natureza.

D. Desempenho dos casos em relação aos *Standards* de Precisão (PC)

Dentre os 12 *Standards* de Precisão, os casos apresentaram desempenho crítico no PC12 – *Meta-avaliação*. Nenhum dos cinco casos cumpriu nenhum dos seis *checkpoints* propostos. Apenas um deles menciona a importância da meta-avaliação durante o processo avaliativo, mas não foi feita uma proposta formal que atendesse aos *checkpoints* propostos.

O desempenho dos casos no PC6 – *Informações confiáveis* (72%) e no PC9 – *Análise de informações qualitativas* (70%) foi prejudicado devido à ocorrência de

dois *checkpoints* críticos em cada *Standard*, sendo que os dois estão relacionados à comunicação de limitações metodológicas do processo avaliativo. No primeiro caso (PC6), o *checkpoint* relaciona-se à necessidade do avaliador de comunicar as limitações sobre as fontes de informação e análise de inferência e, no segundo (PC9), de informar fatores que influenciaram as características de análise. Essas fragilidades reforçam o baixo desempenho dos *checkpoints* do PR5 – *Levantamento completo e justo*, especialmente naquele relacionado à comunicação nos relatórios sobre as limitações dos estudos avaliativos.

Tabela 5 – Desempenho geral dos casos nos *Standards*

STANDARDS	Totais *	% **
STANDARDS DE UTILIDADE	159	76%
U1 Identificação dos interessados	15	50%
U2 Credibilidade do avaliador	28,5	95%
U3 Escopo e seleção das informações	28,5	95%
U4 Identificação de valores	26,5	88%
U5 Clareza dos relatórios	24	80%
U6 Entrega a tempo e disseminação de relatórios	16,5	55%
U7 Impacto da avaliação	20,0	67%
STANDARDS DE VIABILIDADE	73	81%
V1 Procedimentos práticos	26,5	88%
V2 Viabilidade política	24,5	82%
V3 Custo-efetividade	22,0	73%
STANDARDS DE PROPRIEDADE	195,5	82%
PR1 Orientação para o serviço	27	96%
PR2 Acordos formais	20	67%
PR3 Direitos dos indivíduos	29,5	98%
PR4 Relações humanas	28,0	93%
PR5 Avaliação completa e justa	22,0	73%
PR6 Divulgação dos resultados	25,0	83%
PR7 Conflito de interesses	23,5	78%
PR8 Responsabilidade fiscal	20,5	68%
STANDARDS DE PRECISÃO	278	77%
PC1 Documentação do Programa	25,0	83%
PC2 Análise de contexto	27,0	90%
PC3 Descrição de propósitos e procedimentos	25,0	83%
PC4 Fontes de informações defensáveis	26,0	87%
PC5 Informações válidas	28,0	93%
PC6 Informações confiáveis	21,5	72%
PC7 Informação sistemática	28,5	95%
PC8 Análise de informações quantitativas	23,0	77%
PC9 Análise de informações qualitativas	21,0	70%
PC10 Conclusões justificáveis	27,0	90%
PC11 Imparcialidade de relatórios	26,0	87%
PC12 Meta-avaliação	0,0	0%

* Soma total dos *checkpoints* totalmente cumpridos e cumpridos parcialmente pelos cinco casos.

** Percentual de *checkpoints* cumpridos somados aos cumpridos parcialmente pelos cinco casos

Fonte: Autor

5.2.2 Análise de performance dos casos segundo proposta de Stufflebeam

A análise proposta por Stufflebeam (2000) refere-se a casos individuais. Prevê a compilação e soma do número de *checkpoints* cumpridos em cada *Standard* que resulta em um **produto** “A”. Esse produto é alocado em intervalos e multiplicado por fatores propostos pelo autor (até 6 X 4; entre 5 e 5,9 X 3; entre 4 e 4,9 X 2 e entre 2 e 3,9 X 1), e os produtos dessas multiplicações são somados, resultando em uma **pontuação** que se enquadra em um dos intervalos que correspondem aos conceitos: excelente, muito bom, bom, ruim e pobre (ver memória de cálculo no anexo B). Para a análise do grupo, foi considerada a mesma lógica proposta, utilizando-se as médias do cumprimento dos *checkpoints* em cada um dos *Standards*. As médias foram multiplicadas por fatores propostos no modelo. Esses intervalos vem a corresponder a um percentual que, por sua vez, corresponde a um conceito (tabela 6).

Os casos foram analisados individualmente utilizando-se os mesmos parâmetros supracitados. Como não é objetivo deste trabalho tecer considerações sobre o desempenho comparado nos casos, eles estão classificados em ordem decrescente de desempenho segundo os critérios propostos, sem sua identificação (figura 4).

Tabela 6 – Performance do conjunto de casos nos *Standards*

Utilidade			Pontuação
			17
26 – 28	(93%)	Excelente	
19 – 25	(68%)	Muito bom	
14 – 18	(50%)	Bom	X
7 – 13	(25%)	Ruim	
0 – 5	(0%)	Pobre	
Viabilidade			Pontuação
			7
11 – 12	(93%)	Excelente	
8 – 10	(68%)	Muito bom	
6 – 7	(50%)	Bom	X
3 – 5	(25%)	Ruim	
0 – 2	(0%)	Pobre	
Propriedade			Pontuação
			18
30 – 32	(93%)	Excelente	
22 – 29	(68%)	Muito bom	
16 – 21	(50%)	Bom	X
8 – 15	(25%)	Ruim	
0 – 7	(0%)	Pobre	
Precisão			Pontuação
			31
45 – 48	(93%)	Excelente	
33 – 44	(68%)	Muito bom	
24 – 32	(50%)	Bom	X
12 – 23	(25%)	Ruim	
0 – 11	(0%)	Pobre	

Fonte: Autor

É possível observar que em três dos quatro conjuntos de *Standards* os casos estiveram muito próximos de uma performance “muito boa”. Nos de Viabilidade o valor máximo para ser considerado “Bom” era 7, igual ao valor atingido. Para os *Standards* de Utilidade e Precisão seriam necessários mais três pontos para a conquista pelo conjunto de casos do conceito “Muito Bom”.

Com relação ao desempenho individual, um dos casos destacou-se em três conjuntos de *Standards* (Utilidade, Viabilidade e Propriedade), ficando os de Precisão apenas um ponto atrás do melhores desempenhos neste quesito.

Figura 4 - Desempenho individual dos casos segundo análise de Stufflebeam

	Pobre	Ruim	Bom	Muito Bom	Excelente	Pontuação
Caso A						
Utilidade						23
Viabilidade						11
Propriedade						23
Precisão						37
Caso B						
Utilidade						16
Viabilidade						6
Propriedade						22
Precisão						37
Caso C						
Utilidade						15
Viabilidade						8
Propriedade						18
Precisão						38
Caso D						
Utilidade						15
Viabilidade						6
Propriedade						23
Precisão						25
Caso E						
Utilidade						16
Viabilidade						7
Propriedade						17
Precisão						33

Fonte: Autor

5.2.3 *Standards* críticos

Visando a preparar uma síntese do resultado da meta-avaliação para ser utilizada na segunda parte da pesquisa, foram identificadas situações de convergência quanto ao não atendimento pelos casos dos *checkpoints* (total ou parcial) relacionados aos 30 *Standards*. A análise de convergência originou uma categorização denominada “*Standards* Críticos” (apresentados na tabela 7) para identificar as situações onde ocorreram três ou mais dessas situações.

Os 17 *Standards* Críticos foram utilizados para a formulação de roteiros de entrevistas, com perguntas abertas e fechadas, estruturados para gerentes e avaliadores (Anexos C e D), e cumpriram três propósitos: 1) Levantar informações complementares sobre os *Standards* Críticos para possibilitar maior entendimento dos fenômenos relacionados aos aspectos por eles abordados; 2) Possibilitar aos entrevistados um maior entendimento sobre os *Standards* e sua aplicação, habilitando-os para emissão de juízo de valor sobre a aplicabilidade desses padrões na realidade brasileira; e, 3) Verificar o grau de concordância com o proposto pelos *Standards*, bem como sugestões para sua adequação à realidade brasileira.

Tabela 7 - Standards Críticos

<i>Standards críticos*</i>	<i>Checkpoints</i> NC/CP
<i>STANDARDS DE UTILIDADE</i>	
U1 Identificação dos interessados	3
U5 Clareza dos relatórios	2
U6 Entrega a tempo e disseminação de relatórios	3
U7 Impacto da avaliação	4
<i>STANDARDS DE VIABILIDADE</i>	
V2 Viabilidade política	2
V3 Custo-efetividade	4
<i>STANDARDS DE PROPRIEDADE</i>	
PR2 Acordos formais	2
PR5 Avaliação completa e justa	2
PR6 Divulgação dos resultados	1
PR7 Conflito de interesses	1
PR8 Responsabilidade fiscal	1
<i>STANDARDS DE PRECISÃO</i>	
PC2 Análise do contexto	1
PC4 Fontes de informações defensáveis	1
PC6 Informações válidas	1
PC9 Análise de informações qualitativas	1
PC11 Imparcialidade de relatórios	1
PC12 Meta-avaliação	6

* três ou mais situações de não cumprimento (NC) ou cumprimento parcial (CP) dos *checkpoints* pelos 5 casos

Fonte: Autor

5.3 SEGUNDA ETAPA DA PESQUISA - POSICIONAMENTO EM RELAÇÃO AOS *STANDARDS*

Nesta etapa são respondidas as questões cruciais da pesquisa. Os dados analisados referem-se a duas rodadas de coleta: 1) entrevistas realizadas com os cinco gerentes e cinco avaliadores, e 2) preenchimento de questionários complementares pelos mesmos entrevistados. Buscou-se aprofundar e confirmar informações geradas a partir da meta-avaliação, discutir o desempenho dos casos nos *Standards* e levantar junto aos entrevistados seu posicionamento quanto à aplicabilidade desses *Standards* na realidade brasileira. Foi realizada a análise qualitativa dos dados gerados pelo conjunto de entrevistas, procedendo-se, enquanto etapa da pesquisa, ao que Yin (2001) denomina de “análise cruzada dos casos”.

Questão 2 - Como os gerentes e avaliadores se posicionam em relação aos *Standards*?

Questão 3 - Quais dos *Standards* são condizentes e aplicáveis, considerando-se o atual estado da arte da avaliação de programas e projetos sociais desenvolvidos por organizações que fazem investimento social privado?

5.3.1 Uso e utilidade dos *Standards*

Na primeira parte da entrevista com os gerentes e avaliadores, procurou-se identificar o grau de familiaridade com os *Standards*. Apenas um dos entrevistados informou conhecê-los e tê-los utilizado na avaliação realizada. O avaliador de um outro caso informou conhecê-los mas não fez uso dos mesmos no caso em estudo. Após uma breve apresentação do processo de construção dos *Standards* e dos propósitos de sua aplicação, solicitou-se a opinião dos entrevistados sobre a necessidade de se discutir e desenvolver um conjunto de *Standards* aplicáveis à realidade brasileira. Os respondentes foram unânimes no reconhecimento dessa necessidade e da sua importância e também foram apontadas algumas preocupações e condições para que um conjunto de *Standards* seja realmente útil:

...“acredito que uma linha de trabalho, ou alguma coisa que pautе o trabalho seja realmente interessante sim”. Gerente 1

...“a gente não tem um padrão para todos e isso seria bastante útil”. Gerente 2

...“interessante eu acho que seria. Só que é difícil chegar numa coisa boa... que não fique assim muito burocrática, e que não perca a dimensão de cada caso e de cada contexto que está sendo trabalhado”. Gerente 3

...“uma primeira contribuição seria na própria intencionalidade da criação dos *Standards*, que é de orientar os gerentes na escolha, na proposta de avaliação. Acho que isso é muito bom, penso também que do ponto de vista dos avaliadores é interessante, desde que isso não se configure (em) amarras”. Avaliador 1

...“eu acho que eles têm que ser flexibilizados... esta proposta tem ingenuamente uma visão positivista”. Avaliador 2

As preocupações foram expressas mesmo com apresentação das recomendações feitas pelo *The Joint Committee of Standards for Educational Evaluation*:

“Os Standards são princípios orientadores e não regras mecânicas. Eles contém orientações para evitar eventuais erros e se fundamentam em práticas generalizadamente aceitas, além de propor diretrizes que refletem as best-practices em avaliação de programas atualmente”⁴³

As análises feitas pelos entrevistados sobre os *Standards* individuais à luz de sua prática e do contexto em que ocorreram as avaliações apontaram que, apesar de que a proposição de uso dos *Standards* pelos seus criadores não seja rígida, a formulação de alguns deles sugere certa rigidez. Isto afeta sua aplicabilidade à realidade brasileira.

⁴³ *The Program Evaluation Standards*, p.8

5.3.2 Posicionamento dos entrevistados em relação aos *Standards*

Os dez entrevistados posicionaram-se com um alto grau de concordância em relação à aplicabilidade da maioria dos 30 *Standards*, como se verifica na tabela 9. Três *Standards* obtiveram 100% do grau máximo de concordância, 11 obtiveram 90%, cinco obtiveram 80% e seis obtiveram 70%. Em termos numéricos, esses dados demonstram que 21 dos 30 *Standards* receberam de pelo menos sete dos dez entrevistados um posicionamento de total concordância sobre sua aplicabilidade no contexto brasileiro.

Esse referencial numérico é útil para a definição de um critério analítico sobre o tópico central da questão três da pesquisa. Os cinco *Standards* que apresentaram menor grau de concordância (60% e 50%) e um outro que, apesar de estar entre o grupo de concordância igual ou superior a 70% recebeu opiniões de discordância sobre sua aplicabilidade, serão alvo de discussão mais aprofundada com base nas análises de possibilidades de aplicação, críticas e sugestões de adaptações para uma melhor adequação à nossa realidade. (Ver tabela 8.)

Tabela 8 – Percentuais de concordância com relação aos Standards

Concordância						Discordância					
Total		Muito		Pouco		Pouco		Muito		Total	
n	%	n	%	n	%	n	%	n	%	n	%
3	100	1	40%	2	30%	1	20%	1	10%	-	-
11	90%	2	30%	2	20%	2	10%				
5	80%	7	20%	6	10%						
6	70%	13	10%								
3	60%										
2	50%										

n = número de *Standards*

Fonte: Autor

Na tabela 9 são apresentados os 30 *Standards* em ordem decrescente de concordância quanto à sua aplicabilidade. Dentre os 14 que obtiveram maior aceitação (grau de concordância muito, 100% - n = 3 e 90% - n = 11) encontram-se apenas quatro do conjunto definido como crítico na primeira etapa. Três dos sete *Standards* de utilidade (U1, U4 e U6) tiveram menor aceitação (entre 50% e 60%), e dois de propriedade, apesar de bom grau de concordância máxima (90% e 70%), receberam posicionamento de discordância por parte de um pequeno percentual de entrevistados (10% e 20%). Dois *Standards* de precisão (PC4 e PC12) foram alvo de menor concordância (60%), sendo que o PC12, identificado como o mais crítico na etapa anterior, recebeu percentuais de menor concordância e de discordância em maior medida que os outros 29 *Standards*.

Tabela 9 – Grau de concordância/ discordância dos Standards

Standards	Grau de concordância			Grau de discordância		
	total	muito	pouco	pouco	muito	total
U2 Credibilidade do avaliador	100%					
U3 Escopo e seleção de informações	100%					
PR3 Direitos dos indivíduos	100%					
*U5 Clareza dos relatórios	90%	10%				
*V3 Custo-efetividade	90%	10%				
PR4 Relações humanas	90%	10%				
PC1 Documentação do Programa	90%	10%				
PC3 Descrição de propósitos e procedimentos	90%	10%				
PC5 Informações válidas	90%	10%				
PC7 Informação sistemática	90%	10%				
PC8 Análise de informações quantitativas	90%	10%				
PC10 Conclusões justificáveis	90%	10%				
*PC9 Análise de informações qualitativas	90%		10%			
*PR8 Responsabilidade fiscal	90%			10%		
PR1 Orientação para o serviço	80%	20%				
*PR2 Acordos formais	80%	20%				
*PC4 Fontes de informação defensáveis	80%	20%				
*PC2 Análise do contexto	80%	20%				
*PR5 Avaliação completa e justa	80%	10%	10%			
*PR7 Conflito de interesses	80%	10%	10%			
*PC6 Informações confiáveis	70%	30%				
*PC11 Imparcialidade de relatórios	70%	30%				
V1 Procedimentos práticos	70%	20%	10%			
*U7 Impacto da avaliação	70%	10%	20%			
*PR6 Divulgação dos resultados	70%		10%	20%		
*V2 Viabilidade política	60%	20%	20%			
U4 Identificação de valores	60%	20%	10%	10%		
*PC12 Meta-avaliação	60%	10%	20%		10%	
*U1 Identificação dos interessados	50%	40%	10%			
*U6 Entrega a tempo e disseminação de relatórios	50%	10%	30%	10%		

* Standards críticos

Fonte: Autor

5.3.3 Análise cruzada de casos e recomendações quanto aos *Standards* com menor grau de concordância

A. *Standards* de Utilidade

U1 - Identificação dos interessados: Pessoas envolvidas ou afetadas pela avaliação devem ser identificadas, de forma que suas necessidades possam ser atendidas.

A diretriz proposta por este *Standard* é que o avaliador deve envidar esforços para envolver o maior número de interessados, que são de alguma forma afetados e são potenciais usuários dos achados da avaliação, desde a sua fase de planejamento. Pressupõe-se desta forma aumentar a chance da avaliação produzir informações úteis para um maior número de usuários do Programa (conselhos diretores, gerentes patrocinadores, legisladores, educadores, etc.).

Práticas constatadas

Em quatro dos cinco casos estudados as avaliações atenderam especificamente aos interesses dos contratantes. Os avaliadores identificaram esses interesses e formularam as perguntas, objetivos ou eixos estruturantes da avaliação. Apenas em um dos casos, os avaliadores ouviram os contratantes da avaliação e propuseram o encaminhamento de um conjunto de perguntas pré-formuladas para

serem discutidas e negociadas, segundo as necessidades de coordenadores e educadores que operavam diferentes núcleos do Programa.

Apesar de, na maioria dos casos, a relação estar restrita ao avaliador e contratante, observou-se o uso de estratégias consistentes de envolvimento de instâncias importantes na tomada de decisão bem como o uso das informações por dois casos, conforme observa-se no depoimento abaixo:

... “minha experiência tem mostrado que, apesar de... sempre trabalhar com a preocupação da avaliação externa, ... o uso do resultado do produto das avaliações está diretamente relacionado ao quanto eu fui capaz, desde o princípio, de envolver os interessados.” Avaliador

Perguntados sobre as eventuais vantagens que o envolvimento de outros interessados traria, constatou-se posições convergentes e divergentes sobre vantagens deste procedimento. Parte dos gerentes (n = 4) entendeu que seria interessante o envolvimento de outros interessados. As vantagens desse envolvimento estariam relacionadas a um enriquecimento da proposta da avaliação pela inclusão da perspectiva daqueles que estão diretamente envolvidos na ação. Somente um dos gerentes entrevistados divergiu dessa possibilidade, chamando a atenção para a dificuldade de conciliação de interesses, já que para o financiador o importante é a imagem, enquanto que para a equipe técnica o resultado na aprendizagem é o que importa. O mesmo gerente chamou atenção para a dificuldade de envolver diferentes interessados de programas amplos que se desenvolvem em diferentes regiões do país.

Dois avaliadores apresentaram distintas visões sobre a inclusão dos interessados:

...“contrariamente às clássicas avaliações externas onde os avaliadores em geral fazem o plano, o relatório e entregam... eu aposto e procuro conduzir a minha ação nesta direção (envolvendo interessados). Eu costumo dizer que eu não tenho medo absolutamente de qualquer contaminação pelo fato desse envolvimento com os demandantes, ... porque isso tecnicamente se controla ...” Avaliador 1

...“eu acho que na minha experiência de avaliação participativa a participação não pode ser um fim em si mesma; ela é um instrumento que me ajuda a aprofundar o conhecimento das minhas hipóteses; então, eu sou um avaliador que é rodeado por hipóteses. Se alguns interessados, mesmo importantes, não sejam pessoas que possam usar ou aprofundar meu conhecimento das hipóteses, eu não os envolveria diretamente na avaliação, porque eu quero diferenciar o que é interessado para o uso da avaliação, do interessado para a implementação do projeto”. Avaliador 2

O Avaliador 1 não vê restrições na inclusão de interessados. Ele não entende que haja riscos de “contaminação” pela participação dos interessados no processo avaliativo. O Avaliador 2 sugere um critério de inclusão de interessados na avaliação, sugerindo o envolvimento daqueles que farão uso da avaliação. Considerando-se a percepção dos gerentes e avaliadores, nota-se que o questionamento com relação à aplicabilidade dos *Standards* está relacionado ao desafio de se conciliar interesses e definir critérios para a participação de interessados desde o início do processo de planejamento da avaliação. Patton (1996) define como *personal factor* a relação uso dos achados da avaliação/envolvimento de interessados. Para o autor, o impacto de uma avaliação está diretamente relacionado à presença e à participação de pessoas que estão estreitamente envolvidas e interessadas em seus resultados. A ausência dessas pessoas desde a concepção da avaliação reduz as chances do uso das informações geradas por ela na tomada de decisão. O autor chama de usuários intencionais todos os interessados que buscam informações para fazer seus julgamentos sobre o Programa. Incluem-se entre esses

usuários tomadores de decisão (financiadores, definidores de políticas públicas, gerentes de programas), participantes operacionais de programas e usuários.

O padrão observado de envolvimento de interessados em quatro dos casos, centrado na relação entre avaliador e cliente, e o direcionamento do foco da avaliação apenas para o interesse do cliente, representa o risco do *gerencialismo* apontado por Guba & Lincoln. A avaliação realizada apenas com base nos parâmetros estabelecidos por gerentes ou por outros representantes da organização patrocinadora é pouco inclusiva, por atender a interesses unilaterais de uma relação dita “entre parceiros”. O eventual interesse em apenas comunicar as ações sociais positivas de uma organização pode levar o contratante de uma avaliação a um estabelecimento de acordos eticamente questionáveis. Esses interesses remetem a modelos definidos por Stufflebeam (2001) como pseudo-avaliativos.

A proposta de Stake do envolvimento dos interessados de maneira responsiva na avaliação, ou seja, o avaliador atuando orientado única e exclusivamente pelas demandas desses atores, somadas à abordagem inclusiva, proposta por House & Howe; e ainda, o modelo de inclusão e negociação proposto por Guba & Lincoln, apontam para aspectos ausentes em quatro dos casos. Percebeu-se que, embora os avaliadores valorizem o maior envolvimento dos interessados, não houve por parte dos contratantes a percepção, valorização e conseqüente solicitação para a inclusão no processo avaliativo de outros interesses que não os seus próprios.

Recomendação

Apesar de não haver concordância total quanto à aplicabilidade do U1 no contexto brasileiro, há riscos quanto à não adoção de práticas inclusivas, conforme indicam as argumentações dos autores supracitados. Portanto, recomenda-se a

manutenção deste *Standard* tal qual ele foi proposto. O envolvimento de um maior número de interessados desde a fase do planejamento das avaliações deve ser algo a ser perseguido pelos avaliadores brasileiros. Os limites do envolvimento, entretanto, devem ser alvo de discussões mais aprofundadas, já que avaliações que envolvem um grande número de interessados podem consumir muito tempo e recursos no processo de consulta e consenso sobre os interesses a serem atendidos.

U4 - Identificação de valores: As perspectivas, os procedimentos e as justificativas usados para interpretar os resultados da avaliação devem ser cuidadosamente descritos, de forma que as bases para julgamentos de valores fiquem bem claras.

O *Standard* U4 trata da raiz do termo avaliação, que diz respeito à determinação da valia ou do valor de algo. Em um contexto de envolvimento de múltiplos interessados, é necessário determinar e justificar *a priori* os procedimentos escolhidos, bem como as bases de julgamento a serem utilizadas para evitar más interpretações e questionamentos dos achados do processo avaliativo. Por exemplo, a escolha dos objetivos como base de julgamento ou a opção por uma abordagem quase-experimental devem ser detalhadamente descritas e justificadas.

Práticas Constatadas

Nenhum dos casos deixou de cumprir em alguma medida este *Standard*. Somente em um deles o procedimento foi descrito de maneira superficial. Devido ao bom desempenho dos casos na primeira etapa da pesquisa (*Checklist*) este *Standard* não foi alvo de aprofundamento na segunda etapa. Um bom exemplo da preocupação

na definição das bases de julgamento é aquele no qual o Programa buscava apoiar projetos de melhoria da qualidade da educação. Houve um grande empenho do avaliador em definir parâmetros de julgamento dos resultados dos projetos com base no entendimento dos membros do Comitê Técnico que fazia a seleção dos projetos. Esse esforço ocorreu como parte do planejamento da avaliação. Foram definidas oito dimensões para servirem de base de julgamento para o impacto dos projetos na qualidade da educação.

Um outro caso partiu de um grupo de perguntas estabelecidas conjuntamente com o demandante da avaliação, para levantar indicadores de campo por meio de entrevistas individuais e grupos-focais. A análise de dados orientou não só a definição de indicadores como também a revisão de perguntas orientadoras da avaliação. Neste caso, adotou-se declaradamente a abordagem de Guba & Lincoln.

Apenas um dos respondentes que expressou a concordância parcial justificou sua posição, questionando a viabilidade de se estabelecerem critérios *a priori* em todas as situações. Segundo o entrevistado, o avaliador necessita, em algumas situações, verificar hipóteses, por não dispor de informações suficientes para definir critérios. A preocupação expressa pelo entrevistado pode ser válida quando se possui pouca documentação ou em situações em que houve substituição de informantes-chave sobre determinado programa. Do contrário, a existência de fontes de informações é suficiente para o avaliador formular bases de julgamento *a priori*.

A proposição de hipóteses é um procedimento típico de pesquisa e volta-se à conceituação e ao entendimento de fenômenos, enquanto que a avaliação presta-se a julgar o mérito e a relevância de determinado objeto, buscando descrever os motivos pelos quais este “objeto” obteve ou não sucesso. Worthen, Sanders & Fitzpatrick

(1987, p. 30) expressam de forma bastante clara a distinção entre pesquisa e avaliação:

“A avaliação procura analisar a utilidade social diretamente. A pesquisa pode evidenciar indiretamente a utilidade social, na medida em que a verificabilidade empírica do fenômeno geral e a consistência lógica poderão eventualmente ser socialmente úteis. Julgar o valor é questão ‘sine qua non’ da avaliação. Para discriminar um avaliador de um pesquisador pode-se perguntar se a investigação conduzida seria considerada um fracasso caso não produzisse nenhum dado sobre a utilidade do objeto estudado. Um pesquisador, respondendo estritamente como pesquisador, irá provavelmente responder que não.”

Recomendação

A partir da verificação das práticas e da concordância total de 60% dos entrevistados, aliadas à fragilidade da argumentação de concordância, é razoável a aceitação do *Standard* U4 como aplicável sem alterações, considerando-se as limitações desse estudo.

U6 – Entrega a tempo e disseminação de relatórios: Achados significativos e relatórios de avaliação devem ser disseminados junto aos usuários intencionais da avaliação, para serem utilizados no momento adequado.

O U6 trata de diferentes aspectos referentes aos relatórios e à disseminação. O mais importante deles é o fornecimento de informações de qualidade em tempo para que os usuários intencionais façam o melhor uso delas. Outro aspecto abordado diz respeito à necessidade do avaliador desenvolver, em conjunto com o cliente, formas de comunicação adequadas aos diferentes públicos. Como complemento a esse

Standard, o PR 2 – *Acordos formais*, sugere que o avaliador defina no contrato os formatos de relatórios, as audiências e os prazos.

Práticas Constatadas

Quatro dos cinco casos apresentaram problemas com prazos. Em apenas um desses quatro, tanto o gerente quanto o avaliador informaram que o atraso foi negociado. Ocorreu que, desde o início do processo, tentou-se estabelecer um prazo inviável para a dimensão do trabalho contratado. Nos outros três casos nos quais o problema ocorreu, os gerentes informaram alguns incômodos com os atrasos:

...“no caso de algumas avaliações que precisam de correção, de redação ... você estar revisando ... compromete inclusive até o próprio processo de avaliação, porque na realidade a decisão não sendo tomada naquele momento, você acaba protelando por meses, e no caso de ano letivo, às vezes até perde uma oportunidade... lá na ponta.” Gerente 1

...“quando o relatório já chega está em cima da hora do pessoal fazer o planejamento do próximo ano, fica pouco tempo para poder digerir o relatório, e a gente ter um tempo para poder falar: então, olha, vamos nos assegurar que as recomendações dos avaliadores estão realmente incorporadas no plano de ação.” Gerente 2

...“você tem um financiador, que está esperando uma resposta e você... chegar afobado com a resposta é uma coisa, agora você chegar sossegado, tranqüilo, com tudo revisto, é outra coisa.” Gerente 3

Além dos prazos, outros dois aspectos foram identificados como frágeis. O primeiro diz respeito aos diferentes formatos de relatórios, e o segundo à disseminação a diferentes públicos. O tópico da disseminação será tratado com maior profundidade na análise e discussão do PR6 – *Divulgação dos resultados*. Com relação a formatos, nos cinco casos foram constatados procedimentos similares,

foram produzidos relatórios parciais e finais, um sumário executivo e foram feitas reuniões de apresentação para os contratantes das avaliações. A fragilidade no aspecto formato de relatórios constatada na primeira etapa da pesquisa refere-se ao não preparo de *releases* pelos avaliadores, para a divulgação dos achados.

Com relação ao posicionamento dos entrevistados no que se refere à concordância com a proposta do *Standard*, o aspecto prazo foi o ponto mais criticado, principalmente pelos avaliadores:

... *“é complicado, porque não acho que o relatório está feito para tomar decisão. Não concordo... quer dizer, não é que eu não concorde, é que eu acho que não possa ser uma regra geral.”* Avaliador 1

... *“produzimos relatórios parciais com informações relevantes nos tempos oportunos. Eu quero dizer que, para mim, tempo oportuno é aquele que eu me comprometi a fazer, isto é, eu cumpro adequadamente nos tempos negociados dentro do projeto.”* Avaliador 2

... *“este imediatismo... em educação nada é tão imediato que porque eu deixei de informar a pessoa vai deixar de tomar a decisão ..., nem no Estados Unidos, quando eles fazem estas avaliações grandes, eles informam também de tão imediato.”* Avaliador 3

Três dos entrevistados questionam a validade do relatório como instrumento de tomada de decisão. Um deles traz a perspectiva dos relatórios parciais e dos prazos negociados.

Um dos gerentes aponta para questões estruturais que impossibilitam o cumprimento de prazos não só da avaliação como dos próprios projetos, tanto por operadores, quanto por financiadores:

... *“idealmente pode ser importante, mas para o concreto, eu vejo que a gente tem atraso em todos os projetos... talvez seja a cultura, as dificuldades de estruturas,*

as dificuldades organizacionais, enfim, a gente, financiador e ONG, ... sempre trabalha no limite das coisas.” Gerente

Um dos riscos apontados pelos formuladores dos *Standards* é a produção e disseminação de informações incompletas e imprecisas para a tomada de decisões. Portanto, um relatório parcial e final, mais do que um mero instrumento burocrático, deve assegurar a qualidade da informação disponibilizada.

Recomendação

As observações dos entrevistados sugerem dois aspectos que podem ser combinados: o primeiro é a necessidade de levar em consideração os traços culturais e o estágio de profissionalização das organizações, que determinam certa dificuldade no cumprimento de prazos pré-estipulados; o segundo é a necessidade de um constante diálogo visando a possibilitar a geração de informações úteis em prazos negociados. O constante diálogo para definir o melhor prazo para a disseminação de informações, sem colocar em risco seu uso em processos decisórios, bem como a qualidade e confiabilidade das informações disseminadas, representa um caminho para o atendimento das duas necessidades expressas.

Portanto, a consideração de aspectos culturais e a necessidade de definição de prazos negociados podem ser inclusos no enunciado do U6 para sua melhor adequação à realidade brasileira.

Sugestão de texto

U6 – Entrega a tempo e disseminação de relatórios: O fornecimento de relatórios intermediários e finais a serem disseminados junto aos usuários

intencionais deve obedecer a prazos negociados, levando-se em consideração os limites de tempo para a utilização das informações a serem disponibilizadas.

B. Standards de Propriedade

***PR6 - Divulgação dos resultados:** os responsáveis pela avaliação devem assegurar que todas as pessoas e instituições afetadas pelo processo avaliativo, e aquelas com direito legal, tenham acesso aos resultados da avaliação em sua totalidade, assim como a possíveis limitações relevantes do processo.*

O *Standard* PR6 trata da divulgação dos achados e da sua total disponibilização para todos os que foram, em alguma medida, afetados pela avaliação. A única exceção da não disponibilização completa das informações se faz quando as informações podem colocar em risco a integridade moral ou o direito de livre expressão dos indivíduos. Esse *Standard* fundamenta-se na premissa de que pessoas ou grupos afetados pela avaliação que não têm contato com os achados da avaliação, não podem melhorar sua atuação, bem como não podem contestar bases de julgamentos que porventura os tenham prejudicado. Neste caso, esses interessados passam a ser vítimas de processos avaliativos mal conduzidos. Assim, o PR6 prevê que todo avaliador deve estar pronto para disponibilizar toda a informação utilizada e gerada pelo processo avaliativo para análises por agentes externos.

Práticas constatadas

Os cinco casos cumpriram parcialmente a proposta do PR6. Em três deles os critérios e estratégias de divulgação foram definidos exclusivamente pelas instituições contratantes da avaliação. Mesmo em um dos casos cuja definição das estratégias de uso dos achados foi realizada em conjunto com os avaliadores, a divulgação dos resultados esteve restrita aos tomadores de decisão e operadores do Programa. Em três outros casos, parte das informações foi utilizada para processos internos de tomada de decisão e posteriormente foram preparadas para a comunicação externa. Em um deles as informações foram divulgadas e discutidas de forma ampla entre os diferentes interessados.

Houve discordância por parte de alguns dos gerentes entrevistados quanto ao público que deve ter acesso à informação completa:

... “eles podem até ter uma visão, vamos dizer, distorcida, porque eles não estão inseridos naquele todo, naquela parte. Então, por isso que não é interessante dar o relatório inteiro para aquelas pessoas. Deve-se mostrar os positivos e os negativos para todo mundo, mas de forma que seja entendido e de utilidade para todo o grupo.” Gerente 1

... “eu acho complicado... nesse sentido de que você vai... o ‘cara’ recebe o documento, ele vai olhar. Às vezes não está nem preocupado, vai entender mal, porque ele não tem uma capacidade de leitura... Ele tem que ser reescrito para cada público, para que tenha sentido. Mas não é esconder informação..... quem se interessar, ou quem se sentiu incomodado pelas perguntas que foram feitas, acho que tem direito.” Gerente 2

A preocupação está centrada na capacidade de interpretação da informação. As falas expressam preocupação com possíveis más interpretações dos achados. Não há, aparentemente, uma preocupação em ocultar ou negar o acesso a informações para os interessados que vierem a solicitá-las.

Dois dos avaliadores entrevistados expressaram preocupação com o entendimento de que o avaliador é o único responsável por levar a cabo o proposto pelo PR6. Comentaram que já vivenciaram vários processos em que a decisão final da divulgação é do contratante, independentemente da sua recomendação de disponibilizar a informação completa para os diferentes interessados no Programa.

Os formuladores dos *Standards* alertam sobre o risco de se divulgar informações tomando-se por base conveniências institucionais ou interesses econômicos. Esse tipo de conduta tem implicações éticas e legais. Aqui novamente é possível incorrer no uso das pseudo-avaliações, desenvolvidas com interesses apenas de comunicar aspectos positivos do programa, encobrindo-se eventuais fragilidades. São indicados erros que devem ser evitados pelos avaliadores para evitar situações de não cumprimento das premissas éticas expressas pelo PR6. Esses alertas corroboram com as preocupações expressas pelos avaliadores quanto à dificuldade de definir juntamente com o cliente o conteúdo a ser divulgado. Um dos erros é delegar ao cliente decisões de censura e edição de informações, mudando o sentido expresso dos relatórios preparados pelos avaliadores; outro, é não estar envolvido no processo de preparação de *releases* para a mídia.

Recomendações

Os questionamentos expressos pelos entrevistados não invalidam a aplicabilidade da diretriz expressa pelo PR6. Parte dos interessados revelam preocupação com a forma de se comunicar e outro grupo centra suas preocupações na co-responsabilidade entre avaliador e contratante quanto ao compromisso de divulgar ou disponibilizar as informações, como proposto nos *Standards*. Neste caso,

cabe uma sugestão de redação para o *Standard* em questão, incluindo o compromisso e a co-responsabilidade na divulgação dos achados.

Sugestão de texto

PR6 - Divulgação de resultados: o contratante e o avaliador devem ser co-responsáveis pela divulgação dos resultados e assegurar que todas as pessoas e instituições afetadas pelo processo avaliativo, e aquelas com direito legal, tenham acesso aos resultados da avaliação em sua totalidade, assim como às possíveis limitações relevantes do processo.

C. Standards de Viabilidade

V2 - Viabilidade política: A avaliação deve ser planejada e conduzida tendo-se clareza antecipada dos diferentes posicionamentos dos vários grupos de interesse. Deve assim conseguir a cooperação de todos e tentar evitar, ou criar mecanismos para contra-atacar, possíveis esforços, de qualquer destes grupos, para dificultar a operacionalização, ou confundir, ou usar de forma inadequada os resultados da avaliação.

O *Standard V2* expressa as preocupações e cuidados com a dimensão política intrínseca em um processo avaliativo. Os interesses de grupos que buscam influenciar políticas ou direcionar recursos para suas ações trazem sérias implicações e pressões sobre os diferentes envolvidos nas avaliações. O *V2* sugere que o

avaliador esteja vigilante para que uma avaliação não sofra manipulações e não venha a perder credibilidade.

Práticas constatadas

Apenas em um dos casos houve procedimentos prévios relacionados a cuidados com a prevenção a eventuais conflitos de interesse ou outras iniciativas que pudessem colocar em risco o processo avaliativo:

Em outro caso, houve um processo de discussão prévia que cumpriu em boa medida a diretriz proposta pelo V2 antes da contratação dos avaliadores. Segundo o entrevistado, os avaliadores contratados deram continuidade ao processo:

... “houve um processo que se iniciou entre financiador e financiado, um longo processo de definição para chegar em que tipo de avaliação a gente gostaria, aí que tipo de avaliação que seria útil para ambos os lados, e que tipo de avaliador a gente estaria selecionando. Aí, quando nós selecionamos os avaliadores, houve uma preocupação explícita, depois que os avaliadores foram selecionados, de estar clareando quais eram os interesse explícitos de cada um, das organizações financiadoras e dos financiados.” Avaliador

Um outro entrevistado apresentou uma situação conflituosa, mas que não foi discutida ou minimizada pelo processo avaliativo:

... ”o único pontinho que a gente tem permanentemente é que obviamente um financiador do mundo privado, ele não é só... ele está aí com um interesse de imagem que ele quer que volte para ele, uma imagem positiva. E isso às vezes se sobrepõe ou acelera, querem que acelerem alguma coisa que é mais lenta se você vir sob o ponto de vista social. Mas isso também é uma tensão...” Gerente

Dois avaliadores expressaram posicionamentos críticos com relação ao proposto pelo V2:

“Deve ter clareza. Eu acho que eu concordo um pouco com isso... no início... eu acho isso muito difícil, acho até que é uma recomendação, porque na verdade da minha experiência, só nesse caso, você vai tendo clareza política no decorrer da avaliação... na minha história de avaliadora acho que isso aqui não dá para fazer, mesmo, não é prático, não é viável.” Avaliador 1

“A minha experiência diz que essa clareza você não consegue de cara, é uma coisa a ser construída. Então eu acho que é muito positivista ... Eu acho que às vezes esses Standards tem uma visão muito linear do que é evolução. Você não controla todas as variáveis de um projeto com atores sociais fortes. Você só consegue controlar com atores sociais fracos. Para isso você não precisa fazer grandes avaliações. Eu acho que avaliações... têm que trazer um pouco mais de aceitação da contradição entre os interesses...” Avaliador 2

As críticas dos avaliadores referem-se ao caráter de predição do enunciado do *Standard*, para conflitos que emergem e são perceptíveis durante o processo avaliativo. O estudo mais aprofundado da proposta do V2 confirma em parte o caráter positivista interpretado pelos entrevistados. Há uma orientação para que o avaliador ausculte o maior número de interessados na fase de planejamento da avaliação para a identificação de conflitos. Propõe-se, também, que se explicite, no contrato, cláusulas sobre o interesse público da avaliação, definindo-se critérios para a edição e a publicação de informações por ela geradas. Entretanto, não são explicitados como diretrizes os procedimentos para minimizar situações conflituosas que possam emergir durante o processo avaliativo. A única orientação neste sentido é que, no caso de emergirem situações políticas que coloquem em risco a avaliação, deve-se suspender o processo.

Recomendação

Para melhor adequação deste *Standard*, recomenda-se alterar o sentido da expressão “deve-se ter clareza” e incluir a dimensão de “lidar com contradições”.

Alguns elementos do sentido literal do texto original também podem ser utilizados, buscando-se um sentido em conteúdo e lingüística mais apropriados.

Sugestão de texto

V2 - Viabilidade política: O planejamento da avaliação deve prever ações para lidar com as diferentes posições dos diferentes grupos de interesse que possam emergir durante o processo avaliativo. As ações devem estar voltadas para a obtenção da cooperação dos diferentes grupos e, também, para salvaguardar a avaliação de eventuais tentativas de determinados grupos de enviesar ou fazer uso inadequado de resultados.

D. Standards de Precisão

PC12 - Meta-avaliação: o processo de desenvolvimento, bem como o possível impacto da avaliação para o programa, devem ser avaliados utilizando os Standards descritos aqui, ou outros, a fim de garantir que o processo transcorra de forma adequada e que, ao final, os principais interessados/afetados pela avaliação possam visualizar com clareza os pontos positivos e as limitações da avaliação.

A proposta do PC12 fundamenta-se na premissa de que a avaliação de programas é um importante recurso para prover informações para a tomada de decisão sobre a eventual continuidade, modificação e suspensão de determinado projeto, serviço ou sistema. É esperado que as informações fornecidas sejam confiáveis para orientar as decisões e para que não se chegue a conclusões

equivocadas, incorrendo-se no risco de cometer-se injustiças com o programa, seus condutores e beneficiários. Recomenda-se o procedimento da meta-avaliação formativa e somativa por avaliadores externos, para evitar erros durante o processo avaliativo e para, ao final do programa, poder-se aferir o mérito e a relevância da avaliação realizada.

Práticas constatadas

Apenas em um dos casos verificou-se a inclusão da intenção de meta-avaliações formativas e somativas com base nos *Standards*. Nos relatórios analisados na primeira fase da pesquisa não foram encontradas evidências de que os procedimentos propostos tenham sido implementados. Em entrevista realizada na segunda etapa da pesquisa, o gerente informou que estava em curso uma discussão sobre a realização da meta-avaliação. Embora em nenhum dos outros casos tenha ficado evidente a intenção de se fazer meta-avaliação de maneira formal, dois dos entrevistados mencionaram cuidados que são tomados para assegurar a qualidade das informações geradas, independentemente da utilização de procedimentos e parâmetros formais:

... “tenho a impressão que isso não é uma regra explícita, mas ela é implícita no fazer da avaliação dos brasileiros também, porque na hora que você pega o relatório avaliativo e entrega finalmente, existe no mínimo uma, duas ou três oficinas com os principais interlocutores onde você discute com base no relatório de avaliação, portanto, com base no todo, como foi o processo de avaliação e a que resultados chegamos; e obviamente aí tem espaço tranquilo para ser avaliado o próprio relatório avaliativo.” Avaliador 1

... “entendo que a meta-avaliação ainda é uma sofisticação para a maioria dos empreendimentos da avaliação. Entretanto, é possível e necessária de ser realizada... Por outro lado, poder-se-ia dizer que ela até exista na atuação dos programas aqui incluídos, mas feita informalmente, sem a sistemática relação com

os Standards de meta-avaliação. Valeria a pena observar essas manifestações.”
Avaliador 2

Houve convergência na visão dos gerentes quanto à importância do PC12 na realidade brasileira. Um deles, apesar de julgar a prática da meta-avaliação importante, divergiu quanto à sua aplicabilidade, devido ao investimento de recursos extras nesta atividade. Dois avaliadores concordaram totalmente sobre a necessidade e aplicabilidade dos PC12. Três outros discordaram, alegando inadequação por questões culturais e de custos. Um deles faz uma análise das questões que determinam os traços culturais não receptivos a esse tipo de prática:

“Eu acho que a gente não vai chegar a isso, eu acho que isso é inviável, especialmente considerando que a cultura de avaliação no Brasil é uma cultura recente, ela vem muito associada à questão neoliberal, que implica em entender o estado avaliador. Por isso ... precisa avaliar, precisa avaliador externo, o terceiro setor precisa de avaliação externa para poder ter financiamento, para poder ter prêmios, etc. As políticas públicas põem essas exigências, inclusive nos financiamentos dos programas. No ministério é uma exigência, se não tiver avaliação não tem financiamento, então isso é o que? É um Estado avaliador fazendo isto. Mas isso tudo é muito recente, é muito recente, é uma cultura que está entrando por conta de uma ideologia.” Avaliador

A ausência da prática de meta-avaliação nos casos relatados, aliada aos aspectos culturais e de custo expressos pelos avaliadores, sugere a não adequação dessa prática como um *Standard* por questões de viabilidade. Faz-se necessário, entretanto, assegurar a credibilidade dos processos avaliativos e dos próprios avaliadores. Na visão de Stufflebeam (2001b), a meta-avaliação é uma obrigação profissional dos avaliadores, já que existem avaliações boas, ruins e medianas. Para o autor, muitos avaliadores podem incorrer em erros como a proposição de critérios de julgamento inadequados, erros metodológicos, custos excessivos, abuso de autoridade, enviesamento de achados e outros. Há, portanto, uma preocupação com

procedimentos e condutas do avaliador remetendo à idéia de um código de ética profissional. No Brasil, não há um grau de maturidade nas práticas avaliativas que fomente procedimentos técnicos e éticos recomendáveis para atuação dos avaliadores. Desta forma, o profissional que realiza avaliações ainda não é reconhecido como “avaliador”; normalmente esses profissionais são “consultores” ou “pesquisadores”.

Recomendação

Dadas as condições culturais e de maturidade do setor, o *Standard* de meta-avaliação tal qual foi proposto deve ser considerado como uma diretriz a ser perseguida pelo setor de investimento social privado, como estratégia de comprovação da qualidade e de credibilidade das avaliações realizadas. Deve-se também buscar instrumentos e estratégias que tornem o procedimento menos oneroso.

Tabela 10 - Resumo das análises e recomendações

STANDARDS	Práticas <i>Checklist</i> (% C + CP)	Concor- dância	Recomen- dação
STANDARDS DE UTILIDADE	76%		
*U1 Identificação dos interessados	50%	50%	Manter
U2 Credibilidade do avaliador	95%	100%	Manter
U3 Escopo e seleção de informações	95%	100%	Manter
U4 Identificação de valores	88%	60%	Manter
*U5 Clareza dos relatórios	80%	90%	Manter
*U6 Entrega a tempo e disseminação de relatórios	55%	50%	Alterar
*U7 Impacto da avaliação	67%	70%	Manter
STANDARDS DE VIABILIDADE	82%		
V1 Procedimentos práticos	88%	70%	Manter
*V2 Viabilidade política	82%	60%	Alterar
*V3 Custo-efetividade	73%	90%	Manter
STANDARDS DE PROPRIEDADE	81%		
PR1 Orientação para o serviço	96%	80%	Manter
*PR2 Acordos formais	67%	80%	Manter
PR3 Direitos dos indivíduos	98%	100%	Manter
PR4 Relações humanas	93%	90%	Manter
*PR5 Avaliação completa e justa	73%	70%	Manter
*PR6 Divulgação dos resultados	83%	70%	Alterar
*PR7 Conflito de interesses	78%	80%	Manter
*PR8 Responsabilidade fiscal	68%	90%	Manter
STANDARDS DE PRECISÃO	77%		
PC1 Documentação do Programa	83%	90%	Manter
*PC2 Análise do contexto	90%	80%	Manter
PC3 Descrição de propósitos e procedimentos	83%	90%	Manter
*PC4 Fontes de informações defensáveis	87%	60%	Manter
PC5 Informações válidas	93%	90%	Manter
*PC6 Informações confiáveis	72%	70%	Manter
PC7 Informação sistemática	95%	90%	Manter
PC8 Análise de informações quantitativas	77%	90%	Manter
*PC9 Análise de informações qualitativas	70%	90%	Manter
PC10 Conclusões justificáveis	90%	90%	Manter
*PC11 Imparcialidade de relatórios	87%	70%	Manter
*PC12 Meta-avaliação	0%	60%	Manter

* *Standards* críticos – três situações de não cumprimento (NC) ou cumprimento parcial (CP) dos *checkpoints* pelos casos.

Fonte: Autor

6 CONCLUSÕES E RECOMENDAÇÕES

Este estudo inicia-se com descrições contextuais da abrangência da ação social do terceiro setor no Brasil e em outros países, sobre a dimensão da contribuição do setor privado a essas iniciativas e faz referência à ausência de sistemáticas de avaliação de programas. Paralelamente a esse contexto, são apresentadas informações sobre as organizações de profissionais que se dedicam a fomentar pesquisas, definir princípios e disseminar conceitos e *best practices* em avaliação de programas e projetos ao redor do mundo. A *American Evaluation Association* foi criada em meados da década de 80 como resultado da maturidade da avaliação enquanto área de conhecimento. Na década de 90, o agrupamento de profissionais em torno de uma agenda para incentivar e aprimorar as práticas avaliativas nos seus países expandiu-se para a Europa, Ásia e América Latina. No Brasil, o processo de democratização, a revisão do papel do Estado e a emergência do terceiro setor criaram um ambiente favorável para a introdução de práticas e discussão de modelos de avaliação que podem ser empregados para avaliar a implementação de políticas públicas e as ações criadas por investimentos sociais privados.

Segundo Falconer (1999), as fundações e os institutos empresariais não são reconhecidos pela literatura internacional como organizações do terceiro setor, pela

sua vinculação aos interesses de seus mantenedores. Neste estudo, entretanto, não se buscou refinar este aspecto, mas focar avaliação de projetos e programas sob a ótica do investidor ou patrocinador das ações sociais de benefício público. As ações sociais alvo das avaliações nos cinco casos estudados são intersetoriais, resultantes de parcerias entre as empresas ou suas fundações, organizações da sociedade civil e instituições públicas.

O estágio de desenvolvimento ou maturidade da grande maioria das empresas que investem em ações sociais, no que se refere às relações de parceria, é ainda filantrópico, segundo o continuum proposto por Austin. Há, entretanto, um pequeno grupo de empresas, institutos e fundações que, além de estabelecer relações integradas com seus beneficiários, utilizam-se de estruturas, organizações e sistemas de gestão sofisticados e organizam-se em grupos associativos para troca de experiências, aprimoramento de práticas e disseminação de suas ações. Neste segundo grupo foi possível identificar os cinco casos tratados nesta dissertação.

A discussão dos *Standards*, considerando o atual cenário do investimento social privado no Brasil, foi bastante apropriada na visão dos entrevistados. Houve tanto da parte dos gerentes quanto dos avaliadores muito boa receptividade sobre a idéia de se discutir e delinear princípios, critérios e diretrizes para orientação do *design* de avaliação de programas e também para avaliar as práticas avaliativas vigentes. A idéia de utilização de *Standards* para avaliação de projetos e programas na realidade brasileira foi aceita, adotando-se a compreensão dada a eles como princípios orientadores. Não é recomendada pelos formuladores e nem foi aceita pelos entrevistados a idéia de padrões ou normas de certificação.

O percurso do estudo do desenvolvimento de modelos e abordagens de avaliação remete claramente a práticas preocupadas em lidar com a complexidade e a pluralidade das sociedades democráticas contemporâneas. A demanda pela inclusão

de interessados proposta pelas abordagens (responsiva de Stake, construtivista de Guba & Lincoln, democrática e deliberativa de House e Howe, e focada na utilização de Patton) não é ideológica e sim técnica. A necessidade de contemplar e acomodar valores de diferentes atores envolvidos nas interações intersetoriais é premente numa sociedade democrática. A idéia dessa interação entre diferentes grupos e atores atende à afirmação de Habermas⁴⁴ de que os homens não podem se comunicar de maneira autêntica em um contexto social repressivo. Segundo o autor, para que isto ocorra, todos os interessados em participar do discurso devem ter condições iguais de fazer e refutar afirmações, interpretações e recomendações. O avaliador como um único juiz e o atendimento a necessidades de um único ator social criam uma situação de desequilíbrio de poder que, por sua vez, pode gerar informações de utilidade limitada para outros atores e de difícil defesa, dependendo dos critérios de julgamento escolhidos.

Fica destacada a compreensão da avaliação como prática distinta da pesquisa acadêmica, ainda que ambas empreguem métodos e técnicas semelhantes. A avaliação não se presta a satisfazer a curiosidade do pesquisador. Ela deve estar orientada para a solução de problemas dos diferentes envolvidos em projetos e programas, reiterando a necessidade de intensa participação dos interessados pela avaliação desde o início do processo avaliativo, para que eles possam fazer uso dos achados e levar a cabo as mudanças recomendadas. Nestes aspectos, a avaliação está mais próxima da pesquisa-ação. De acordo com Thiollent (2002), este tipo de pesquisa pode ser considerado como um procedimento de base empírica realizado de maneira associada a uma ação ou problema de modo cooperativo entre o pesquisador e os envolvidos na situação.

⁴⁴ Habermas, J. *Vorbereitende Bemerkungen zu einer Theorie der kommunikativen Kompetenz*. Apud Freitag, B. & Rounet, 2001.

O estudo dos cinco casos sob a ótica dos *Standards* na primeira etapa da pesquisa forneceu um interessante recorte fotográfico do estado da arte da avaliação de programas do grupo de organizações que faz uso de sistemáticas mais sofisticadas para avaliar as ações que patrocinam e operam. Ficou patente um alinhamento em boa medida das avaliações estudadas em relação aos *Standards*, considerando a escala proposta por Stufflebeam. Nenhum dos cinco casos apresentou desempenho abaixo de “bom”. Quatro dos cinco casos obtiveram pontuação correspondente a “Muito Bom” nos *Standards* de precisão. Embora este não tenha sido o objetivo central do estudo, este resultado da pesquisa demonstra o alto grau de qualificação dos profissionais que conduziram as avaliações estudadas.

As maiores fragilidades constatadas no desempenho dos casos nos *Standards* deu-se em três deles (PC12, U1 e U6). O primeiro refere-se às práticas de meta-avaliação: essa prática pode ainda ser considerada uma sofisticação para o estágio de maturidade da avaliação de projetos e programas sociais no país. O segundo deles refere-se à identificação de interessados: o estudo possibilitou a percepção de que havia um desconhecimento, por parte principalmente dos gerentes, quanto às necessidades e vantagens do envolvimento de interessados desde a fase de planejamento de uma avaliação. O terceiro refere-se aos prazos para a disseminação de informações: é possível concluir que esta seja uma questão crítica e que demande um constante diálogo para se chegar a “prazos negociados”, e para o preparo e disponibilização de informações de qualidade a tempo para assegurar a tomada de decisões.

Em relação ao objetivo principal deste estudo, é possível concluir que 27 *Standards* são aplicáveis tal qual eles foram propostos nos Estados Unidos. Três

deles tiveram o texto modificado para melhor se adequarem ao contexto brasileiro. Obviamente, pela extensão da pesquisa, não é possível concluir que os 30 *Standards* estejam validados para nossa realidade. O que o estudo possibilita é inferir que existe um ambiente favorável à adoção deles para impulsionar a reflexão sobre as práticas de avaliação de programas no segmento do investimento social privado no país.

O contexto sociopolítico e econômico vigente parece oferecer condições favoráveis para a adoção de princípios e critérios construídos sobre uma base de valores sociais como a autonomia e o igualitarismo, que vigoram na sociedade norte-americana. Diferente da realidade de alguns países da Ásia, onde vigoram os valores do conservadorismo e da hierarquia, no Brasil dos últimos anos houve grandes avanços na construção de uma sociedade democrática e igualitária.

Uma vez aceitos os *Standards* como critérios úteis para o julgamento de boas práticas avaliativas – dado o elevado grau de concordância por parte dos entrevistados com relação a essa afirmação –, é possível identificar aspectos frágeis do conjunto de casos que podem ser objeto de reflexões mais aprofundadas em futuras pesquisas. A dimensão da utilidade pouco conhecida e discutida por pesquisadores e gestores exige maior reflexão sobre vários aspectos. Estas constatações dão origem a questões como: Até que ponto as competências acadêmicas de pesquisa são suficientes para a realização de avaliações? Quais os limites e critérios para definir a extensão de participação de interessados para assegurar a utilidade dos resultados da avaliação para um maior número de usuários? A não existência (no universo pesquisado) da prática da meta-avaliação é outro aspecto sobre o qual as seguintes questões podem ser lançadas: Até que ponto existem de fato práticas implícitas de meta-avaliação? Como ocorrem essas práticas? Quão efetivas elas são?

Finalmente, a aplicação de métodos e técnicas de avaliação tem sido discutida nos Estados Unidos como uma área profissional especializada. O perfil de competências exigidas e a responsabilidade que os avaliadores assumem quando estão à frente de processos avaliativos são grandes, o que se observa no conjunto de *Standards* estudados, muito amplos e complexos. A *American Evaluation Association* desenvolveu recentemente um rol de princípios para orientar a conduta ética do avaliador. No Brasil, o surgimento de fóruns de discussão, como a Rede Brasileira de Avaliação, representa um caminho para internalizar e ampliar o debate sobre o tema. Além dessa rede, as universidades brasileiras congregam hoje pesquisadores que realizam avaliações e podem contribuir com reflexões a questões como: Onde se situa a avaliação enquanto campo de conhecimento? Ela é uma disciplina, uma transdisciplina? Qual seria o perfil de competências de um avaliador? E, qual seria o perfil do docente para orientar a modelagem de programas de formação de avaliadores?

6.1 LIMITAÇÕES DA PESQUISA

Esta pesquisa possui limitações inerentes aos estudos preliminares de caráter descritivo e qualitativo. Além das questões subjetivas sempre presentes nas pesquisas das ciências sociais, as seguintes limitações deste estudo em particular podem ser elencadas:

- O restrito conhecimento acumulado no Brasil sobre o tema abordado impossibilitou a elaboração de hipóteses *a priori* para pesquisa. Os estudos no campo da gestão de programas e projetos apoiados pelo investimento social privado e operados por organizações do terceiro setor são recentes, particularmente no Brasil. A restrição de conhecimento nesta área demandou um desenho exploratório de pesquisa.
- A escolha da amostra foi intencional e não probabilística, realizada segundo critérios definidos pelo pesquisador. Embora esta condição não pudesse ser superada pelas restrições óbvias de trabalho de campo, ela implicou em que os resultados deste estudo tenham baixa possibilidade de generalização.
- Os procedimentos metodológicos adotados visaram a assegurar a validade interna da pesquisa. Lançou-se mão de instrumentos quantitativos (*Checklist*), foram realizados pré-testes e utilizou-se a triangulação, que se deu no levantamento de informações comprobatórias da análise de documentos nas entrevistas da segunda

etapa da pesquisa. Além disso, buscou-se um número expressivo de casos para possibilitar um maior espectro de análise. Os atributos de validade interna da pesquisa não são suficientes, entretanto, para superar as limitações de subjetividade dos estudos qualitativos e também não garantem a sua validade externa.

- O escopo da pesquisa foi amplo, na medida em que o estudo abarcou quatro atributos (Utilidade, Viabilidade, Propriedade e Precisão) distribuídos em 30 *Standards*. Esta opção implicou em uma ampla e diversificada gama de dimensões e temas, que impediu maior profundidade nas discussões sobre questões que emergiram a partir da análise dos dados coletados. Essa limitação, porém, é inerente ao caráter exploratório da pesquisa. Pesquisas dessa natureza não buscam respostas para hipóteses e, sim, apontam para questões que possam ser úteis para a formulação de problemas mais específicos para futuras pesquisas. Algumas dessas questões estão expressas nas conclusões do presente estudo.

7 ANEXOS

ANEXO A - INSTRUMENTO PARA META-AVALIAÇÃO⁴⁵

Nome da organização:
Responsável pela avaliação:
Projeto/programa avaliado:
Período de realização da avaliação:
Documentos disponibilizados para a meta-avaliação:
Nome dos entrevistados:

STANDARDS DE UTILIDADE		
U1 Identificação dos interessados		
	Foram definidos com o cliente os interessados mais importantes	
	O avaliador engajou líderes para identificar outros interessados	
	Os interessados ajudaram a identificar outros interessados	
	Foram consultados interessados para a identificação de suas necessidades de informação	
	Foram propostas estratégias para envolvimento dos interessados durante a avaliação no contrato	
	O avaliador manteve o processo de avaliação aberto para a inclusão de novos interessados	
1- cumpriu o item	0,5 - cumpriu parcialmente	0- não cumpriu o item
Total =		

⁴⁵ Adaptado de *Program Evaluation Metaevaluation Checklist (Based on The Program Evaluation Standards)*. Daniel L. STUFFLEBEAM, 2000.

U2 Credibilidade do avaliador		
	Houve contratação de avaliador competente e confiável na opinião do contratante	
	Os interessados demonstraram confiança no avaliador	
	Foi contratado avaliador com capacidade de endereçar as questões colocadas pelos interessados	
	Foi contratado avaliador com competência para lidar com questões relacionadas a gênero, condições socioeconômicas, linguagem e diferenças culturais	
	O avaliador ajudou os interessados a entenderem o processo da avaliação	
	Foram atendidas apropriadamente as críticas e sugestões dos interessados	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

U3 Escopo e seleção de informações		
	O avaliador atendeu as solicitações prioritárias do cliente	
	O avaliador demonstrou flexibilidade para incluir novas perguntas	
	O avaliador entrevistou interessados para determinar suas diferentes perspectivas e necessidades	
	O avaliador ajudou na seleção de perguntas capazes de determinar o mérito do programa	
	O avaliador ajudou na seleção de perguntas capazes de determinar as fragilidades do programa	
	O avaliador facilitou o processo de definição de perguntas prioritárias atuando como negociador na inclusão de perguntas dos interessados	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

U4 Identificação de valores		
	O avaliador considerou todas as fontes relevantes de valores para interpretar os achados da avaliação (necessidades sociais e dos clientes, legislação pertinente, missão institucional e objetivos do programa)	
	O avaliador determinou as partes apropriadas para realizar as interpretações valorais (julgamento) durante o processo avaliativo	
	O avaliador apresentou fontes e formas claras e defensáveis para o julgamento de valores	
	O avaliador distinguiu apropriadamente o peso e a profundidade dos diferentes valores envolvidos	
	O avaliador levou em consideração os valores dos interessados	
	O avaliador ofereceu formas de interpretações alternativas com base em valores conflitantes porém confiáveis	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

U5 Clareza dos relatórios		
	O avaliador apresentou um ou mais relatórios contendo sumário, relatório principal, relatório técnico e apresentação oral	
	O relatório elaborado estava claro, direto, com linguagem compreensível aos interessados	
	O relatório esteve focado nas questões endereçadas no contrato	
	Os achados foram apresentados de forma simples e direta	
	Foram empregadas diferentes mídias para informar diferentes audiências	
	Foram apresentados exemplos para facilitar a compreensão das audiências	
1–	cumpriu o item	0,5 – cumpriu parcialmente
0–	não cumpriu o item	
Total =		

U6 Entrega a tempo e disseminação de relatórios		
	O avaliador procurou, em conjunto com o cliente, identificar, acessar e informar todos os usuários intencionais sobre os achados da avaliação	
	Foram realizadas trocas (discussões) em tempo adequado com a equipe da instituição durante o processo avaliativo	
	Foram realizadas trocas com interessados apropriados (Exemplo: o conselho da instituição, o público-alvo do programa e outros interessados)	
	O relatório final foi entregue no prazo previsto	
	O avaliador desenvolveu <i>releases</i> para divulgação dos achados na mídia	
	Foram utilizadas mídias apropriadas para alcançar e informar diferentes audiências	
1–	cumpriu o item	0,5 – cumpriu parcialmente
0–	não cumpriu o item	
Total =		

U7 Impacto da avaliação		
	O avaliador envolveu os interessados em todo o processo avaliativo apresentando relatórios escritos e (ou) fazendo comunicações verbais	
	O avaliador criou sistemática para prover <i>follow-up</i> e suporte para interpretação e aplicação dos achados	
	O avaliador encorajou os interessados a utilizarem os achados da avaliação, fazendo prognósticos e apontando potenciais usos para os achados	
	O avaliador demonstrou de forma clara como os achados da avaliação poderiam ser incorporados no trabalho da equipe	
	O avaliador suplementou relatórios escritos com constantes comunicações verbais	
	O avaliador conduziu reuniões de <i>feedback</i> para ir além e viabilizar a aplicação dos achados	
1–	cumpriu o item	0,5 – cumpriu parcialmente
0–	não cumpriu o item	
Total =		

STANDARDS DE VIABILIDADE		
V1 Procedimentos práticos		
	O avaliador minimizou fragmentação e excesso de dados	
	Avaliador sugeriu equipe competente e a treinou para o trabalho	
	Foram escolhidos procedimentos que a equipe tinha competência para aplicar	
	Foi proposto um cronograma realista para a realização do trabalho	
	O avaliador envolveu pessoal da equipe do programa para participar da condução da avaliação	
	A avaliação ocorreu sem causar interferências nas atividades de rotina	
1- cumpriu o item	0,5 – cumpriu parcialmente	0- não cumpriu o item
Total =		

V2 Viabilidade política		
	O avaliador antecipou-se às diferenças de posições de diferentes grupos	
	O avaliador preveniu-se contra situações de pressão ou ações deliberadas para impedir a realização da avaliação	
	A avaliador fomentou a cooperação entre os envolvidos	
	O avaliador reportou divergências de pontos de vista	
	Quando possível, o avaliador fez uso das diversas forças políticas para atingir os objetivos da avaliação	
	Rechaçou qualquer tentativa de corrupção à avaliação	
1- cumpriu o item	0,5 – cumpriu parcialmente	0- não cumpriu o item
Total =		

V3 Custo-efetividade		
	O processo avaliativo foi eficiente (cumprimento de prazo, maximização de recursos, trabalho dentro do orçamento)	
	Foi feito uso de recursos do próprio programa durante a avaliação	
	O processo avaliativo produziu informações novas, úteis à tomada de decisão	
	O processo avaliativo fomentou melhorias no programa	
	O processo avaliativo produziu informações sobre o <i>accountability</i> (uso responsável de recursos e prestação de contas às autoridades superiores ligadas ao programa. Exemplo: financiadores e Conselho Diretor)	
	O processo avaliativo propiciou a geração de novos <i>insights</i> sobre o programa	
1- cumpriu o item	0,5 – cumpriu parcialmente	0- não cumpriu o item
Total =		

STANDARDS DE PROPRIEDADE		
PR1 Orientação para o serviço		
	Avaliou os resultados do programa quanto às necessidades do público-alvo e de outros usuários	
	O avaliador assegurou que todos os potenciais usuários de direito do programa fossem contemplados no processo avaliativo	
	Promoveu serviço de excelência	
	Identificou aspectos positivos sobre os quais o programa foi estruturado	
	Identificou aspectos negativos para que fossem corrigidos	
	Apontou com clareza práticas pouco recomendadas ao bom andamento do programa	
1–	0,5 –	0–
cumpriu o item cumpriu parcialmente não cumpriu o item		
Total =		

PR2 Acordos formais		
	Proposta da avaliação e perguntas a serem respondidas pelo processo	
	Definição das audiências da avaliação	
	Definição de formato dos relatórios	
	Estratégias de disseminação dos achados	
	Procedimentos da avaliação, cronograma de atividades	
	Recursos a serem utilizados no processo avaliativo	
1–	0,5 –	0–
cumpriu o item cumpriu parcialmente não cumpriu o item		
Total =		

PR3 Direitos dos indivíduos		
	O avaliador conduziu o processo avaliativo respeitando os direitos civis	
	O avaliador compreendeu os valores dos diferentes participantes	
	O avaliador respeitou a diversidade	
	O avaliador seguiu protocolo pré-estabelecido	
	O avaliador garantiu confidencialidade e anonimato durante o processo avaliativo	
	O avaliador minimizou conseqüências prejudiciais da avaliação	
1–	0,5 –	0–
cumpriu o item cumpriu parcialmente não cumpriu o item		
Total =		

PR4 Relações humanas		
	O avaliador reportou-se a todos os interessados de maneira profissional	
	O avaliador honrou o direito de privacidade dos participantes da avaliação	
	O avaliador honrou os compromissos relativos a prazos	
	O avaliador foi sensível às diferenças culturais e de valores dos diferentes interessados envolvidos no processo avaliativo	
	O avaliador foi imparcial ao endereçar diferentes interessados	
	O avaliador não ignorou ou encobriu eventuais incompetências ou atitudes não éticas, fraudes ou abusos realizados pela equipe do programa	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PR5 Avaliação completa e justa		
	Foram levantados e reportados aspectos fortes e fracos do programa	
	Foram reportadas análises de resultados esperados e não esperados do programa	
	Foi apontado como os pontos fortes do programa podem auxiliar na minimização das fragilidades identificadas na avaliação	
	O avaliador submeteu um relatório preliminar a análise e críticas	
	O avaliador reconheceu as limitações do relatório final	
	O avaliador apresentou no relatório as limitações do processo avaliativo para o julgamento de determinados aspectos do programa	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PR6 Divulgação dos resultados		
	O avaliador identificou claramente as audiências de direito	
	O avaliador reportou pontos relevantes daqueles que suportam e daqueles que criticam o programa	
	O avaliador elaborou relatório balanceado quanto a conclusões e recomendações	
	O avaliador reportou todos os achados por escrito, exceto em circunstâncias em que o registro por escrito não é indicado	
	A elaboração do relatório foi orientada estritamente pelos princípios de abertura e comunicação aberta e direta	
	O avaliador assegurou que os relatórios chegassem a diferentes audiências	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PR7 Conflito de interesses		
	O avaliador identificou potenciais conflitos de interesse logo no início da discussão sobre o processo avaliativo	
	O avaliador envolveu vários avaliadores no processo	
	O avaliador mantém as informações levantadas arquivadas para eventuais análises por auditores externos	
	Se houve viabilidade, o avaliador estabeleceu contrato com o financiador ao invés do beneficiário	
	Se houve viabilidade, o avaliador sugeriu que os relatórios de avaliações internas fossem apresentadas diretamente ao diretor presidente	
	O avaliador envolveu pessoas importantes ao processo avaliativo (Exemplo: membros da equipe do programa) mas que poderiam representar riscos quanto a conflitos de interesse, tomando os devidos cuidados para que esses conflitos não ocorressem	
1–	cumpriu o item	0,5 – cumpriu parcialmente
0–	não cumpriu o item	
Total =		

PR8 Responsabilidade fiscal		
	O avaliador especificou todos os itens de despesa com antecedência	
	O avaliador propôs orçamento flexível para permitir realocações apropriadas para melhor realização da avaliação	
	O avaliador forneceu relatório detalhado sobre a alocação de todas as despesas realizadas	
	O avaliador forneceu relatório detalhado sobre a alocação de tempo utilizado pela equipe de avaliação	
	O avaliador foi criterioso no uso dos recursos realizando apenas gastos necessários ao processo avaliativo	
	Foi incluído no relatório final um informe resumido sobre as despesas realizadas	
1–	cumpriu o item	0,5 – cumpriu parcialmente
0–	não cumpriu o item	
Total =		

STANDARDS DE PRECISÃO		
PC1 Documentação do programa		
	O avaliador coletou e sistematizou descrições sobre intenções do programa de diferentes fontes escritas	
	Foram mantidos registros de várias fontes sobre como o programa opera	
	Foram coletadas e sistematizadas descrições sobre intenções do programa de diferentes interessados	
	Foram analisadas intenções e percepções e foi feita descrição como de fato o programa funcionou	
	Foi solicitada ao cliente e interessados análise das conclusões sobre como o programa de fato funcionou segundo a avaliação	
	Foi produzido relatório técnico (ou capítulo) sobre a operacionalização do programa	
1- cumpriu o item	0,5 – cumpriu parcialmente	0- não cumpriu o item
Total =		

PC2 Análise do contexto		
	Foram descritas as dimensões social, econômica e política de onde ocorreu o programa	
	Foram registradas circunstâncias não usuais ocorridas no programa	
	Foram reportadas influências contextuais que aparentemente influenciaram o programa, que possam ser de interesse daqueles que querem adotá-lo em outro contexto	
	Foram estimados efeitos do contexto nos resultados do programa	
	Foram identificados e descritos quaisquer competidores (projeto, programa) que ocorreram no mesmo período e ambiente do programa	
	Foi descrito como as pessoas afetadas direta ou indiretamente pela existência do programa perceberam a importância e a qualidade do mesmo	
1- cumpriu o item	0,5 – cumpriu parcialmente	0- não cumpriu o item
Total =		

PC3 Descrição de propósitos e procedimentos		
	Houve monitoramento e descrição das mudanças na proposta original ocorridas ao longo do tempo	
	Os procedimentos sofreram modificações ao longo do processo como decorrência das mudanças de propósitos ocorridos	
	Foram registrados os passos sobre como a avaliação foi de fato implementada	
	Foram levantados pontos de convergência e divergência entre os diferentes interessados (incluindo o cliente) sobre os propósitos da avaliação	
	Quando o avaliador realizou a interpretação dos dados, levou em consideração a extensão pela qual os procedimentos previstos foram efetivamente executados	
	Quando viável, o avaliador envolveu outro avaliador para monitorar e avaliar os propósitos e procedimentos da avaliação	
1- cumpriu o item	0,5 – cumpriu parcialmente	0- não cumpriu o item
Total =		

PC4 Fontes de informações defensáveis		
	O avaliador coletou informações prévias pertinentes ao processo avaliativo	
	O avaliador obteve informações de diferentes fontes e métodos de coleta de informação	
	O avaliador registrou e incluiu no relatório as diferentes fontes de informação	
	O avaliador documentou e justificou todas as escolhas feitas em termos de fontes de informação, métodos de coleta e amostra	
	O avaliador incluiu no apêndice do relatório todos os instrumentos utilizados	
	O avaliador registrou e reportou ocorrências que possam ter sido geradoras de vieses	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PC5 Informações válidas		
	O avaliador manteve o foco nas questões-chave definidas	
	O avaliador analisou e reportou qual tipo de informação foi obtido a partir dos procedimentos utilizados	
	Foi documentado como cada procedimento de coleta de informações quantitativas foi escalonado, sistematizado e analisado	
	Foram reportadas e justificadas inferências realizadas individualmente ou em combinação	
	Foi analisado e reportado o nível de compreensão obtido como resultado da utilização dos procedimentos adotados e sua relação com as informações necessárias para responder as perguntas orientadoras da avaliação	
	Foram estabelecidas categorias de significado para temas regulares e recorrentes oriundos da coleta de dados qualitativos	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PC6 Informações confiáveis		
	O avaliador identificou e justificou os tipos e a extensão de confiabilidade pretendidos	
	O avaliador utilizou procedimentos de confiabilidade que no passado demonstraram ser aceitáveis por seus usuários	
	O avaliador reportou fatores que o influenciaram (características de análise, condições em que foi feita a coleta de dados e vieses dos avaliadores)	
	O avaliador checkou e reportou a consistência de escalas, classificações e códigos utilizados	
	O avaliador treinou a equipe, verificou as escalas utilizadas e as análises visando a produzir resultados consistentes	
	Foram realizados pré-testes em novos instrumentos adotados durante o processo avaliativo	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PC7 Informação sistemática		
	Estabeleceram-se protocolos para controle de qualidade das informações geradas pela avaliação	
	Foi verificada a entrada (tabulação) de dados	
	Foram revisadas e verificadas as tabelas de dados gerados por computador ou por outros meios	
	O avaliador sistematizou e controlou o armazenamento (a organização) das informações coletadas	
	O avaliador realizou controle rigoroso do acesso à informação apenas por aquelas pessoas previstas no relatório	
	Os provedores de informações tiveram oportunidade de checar as informações dadas por eles	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PC8 Análise de informações quantitativas		
	O avaliador iniciou a análise por estudos exploratórios para verificar se os dados estavam corretos e para ampliar sua compreensão sobre o conjunto de dados coletados	
	O avaliador reportou as limitações para cada procedimento analítico incluindo os insucessos de confirmação de hipóteses	
	O avaliador empregou múltiplos procedimentos analíticos para checar a consistência e a possibilidade de replicabilidade dos achados	
	O avaliador examinou variabilidade e tendências centrais	
	O avaliador examinou exceções e eventuais correções necessárias	
	O avaliador identificou e analisou interações estatísticas	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PC9 Análise de informações qualitativas		
	Foram definidas as limitações das informações utilizadas	
	Foram definidas categorias analíticas e escolhidos procedimentos de análise e métodos de sumarização apropriados para responder as perguntas avaliativas	
	Foi feita classificação das informações obtidas com base nas categorias de análise adotadas	
	Foi verificada a precisão dos achados através da obtenção de evidências confirmatórias de múltiplas fontes, incluindo os interessados	
	Foram classificadas as informações obtidas de acordo com categorias de validade e confiabilidade	
	O avaliador reportou limitações sobre as fontes de informações, análise e inferências	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PC10 Conclusões justificáveis		
	As conclusões estiveram limitadas aos períodos de tempo, contextos, propostas, questões e atividades do programa	
	Foram reportadas conclusões plausíveis que explicam porque outras conclusões rivais foram rejeitadas	
	Foram citadas informações que suportam as conclusões	
	Foram identificados e reportados os efeitos advindos das ações dos programa	
	Foram adotados procedimentos para evitar más interpretações	
	Quando viável e apropriado, obteve e encaminhou resultados já identificados durante a revisão dos relatório final	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PC11 Imparcialidade de relatórios		
	O avaliador engajou o cliente para determinar os passos para assegurar comunicação imparcial	
	O avaliador salvaguardou o relatório quanto a distorções inadvertidas ou deliberadas	
	Foram divulgadas todas as perspectivas de todos os interessados inclusive daqueles que manifestaram visões opostas ao programa	
	O avaliador obteve avaliadores externos para análise dos relatórios	
	Foram descritos os passos para controlar os vieses	
	O avaliador participou de apresentações públicas para assegurar que não ocorressem distorções dos achados por parte de outros interessados	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

PC12 Meta-avaliação		
	Foram previstos recursos suficientes para a condução de meta-avaliações internas bem como de uma meta-avaliação externa	
	O avaliador definiu previamente <i>Standards</i> a serem utilizados em meta-avaliação a ser conduzida por avaliador externo	
	Foram sistematizadas e armazenadas informações necessárias ao processo de meta-avaliação para atender os <i>Standards</i> de análise predefinidos	
	Foi realizada a contratação de uma meta-avaliação	
	O meta-avaliador avaliou instrumentos, coleta de dados, processamento, análise e relatórios	
	Foram obtidas e reportadas as meta-avaliações formativa e somativa para as audiências de direito	
1– cumpriu o item	0,5 – cumpriu parcialmente	0– não cumpriu o item
Total =		

**ANEXO B - MEMÓRIA DE CÁLCULO DA ANÁLISE DO
CONJUNTO DE CASOS PELA PROPOSTA DE STUFFLEBEAM**

Utilidade (<i>Checkpoints Cumpridos</i>)		A	
Média até 6	x 4	1	4
Média até entre 5 e 5,9	x 3	3	9
Média entre 4 e 4,9	x 2	1	2
Média ente 2 e 3,9	x 1	2	2
Total			17
26 – 28 (93%)	Excelente		
19 – 25 (68%)	Muito bom		
14 – 18 (50%)	Bom	X	
7 – 13 (25%)	Ruim		
0 – 5 (0%)	Pobre		

Viabilidade (<i>Checkpoints Cumpridos</i>)		A	
Média até 6	x 4	0	0
Média até entre 5 e 5,9	x 3	1	3
Média entre 4 e 4,9	x 2	2	4
Média ente 2 e 3,9	x 1	0	0
Total			7
11 – 12 (93%)	Excelente		
8 – 10 (68%)	Muito bom		
6 – 7 (50%)	Bom	X	
3 – 5 (25%)	Ruim		
0 – 2 (0%)	Pobre		

Propriedade (<i>Checkpoints Cumpridos</i>)		A	
Média até 6	x 4	0	0
Média até entre 5 e 5,9	x 3	3	9
Média entre 4 e 4,9	x 2	4	8
Média ente 2 e 3,9	x 1	1	1
Total			18
30 – 32 (93%)	Excelente		
22 – 29 (68%)	Muito bom		
16 – 21 (50%)	Bom	X	
8 – 15 (25%)	Ruim		
0 – 7 (0%)	Pobre		

Precisão (<i>Checkpoints Cumpridos</i>)		A	
Média até 6	x 4	0	0
Média até entre 5 e 5,9	x 3	9	27
Média entre 4 e 4,9	x 2	2	4
Média ente 2 e 3,9	x 1	0	0
Total			31
45 – 48 (93%)	Excelente		
33 – 44 (68%)	Muito bom		
24 – 32 (50%)	Bom	X	
12 – 23 (25%)	Ruim		
0 – 11 (0%)	Pobre		

ANEXO C - ROTEIRO DE ENTREVISTAS - GERENTES

Entrevistado : _____ Data: __/__/__ Cód. _____

Introdução

- Agradecimento pelo apoio ao trabalho.
- Apresentação dos objetivos da pesquisa.
- Descrição das estratégias da pesquisa: 1) 5 casos, 2) *Checklist*, 3) entrevista em profundidade com gerentes e pesquisadores, 4) grupo-foco externo.

Roteiro de Perguntas:

1. Você já conhecia os *Standards* para Avaliação de Programas antes da pesquisa?⁰¹

Sim

Não

Em caso positivo, foi feito uso dos *Standards* para a avaliação estudada nesta pesquisa?⁰²

Sim

Não

Em caso positivo, que tipo de uso foi feito?

Apresentação dos *Standards*:

- Foram criados entre 1989 e 1994 por 15 entidades: AEA, *American Educational Research Association*, pela *American Psychological Association* e pelo *National Council on Measurement in Education*.
- “*Os Standards são princípios orientadores e não uma prescrição rígida. Eles contém orientações para evitar eventuais erros e fundamentam-se em práticas generalizadamente aceitas, além de propor diretrizes que refletem as best-practices em avaliação de programas atualmente*”.
- “*Os Standards devem ser utilizados como orientação para avaliar planos e relatórios... devem ser utilizados como meio para troca de informação sobre a qualidade da avaliação entre o cliente, os avaliadores e os vários stakeholders envolvidos...*”

2. Reação imediata. Você acha que estes 30 *Standards* podem ser úteis para o aprimoramento da prática da avaliação de programas de investimento social privado? Em caso positivo, em que?

Apresentação do resultado do *Checklist* – Quadro com os *Standards* Críticos.

I - UI- Identificação dos interessados: *Pessoas envolvidas ou afetadas pela avaliação devem ser identificadas, de forma que suas necessidades possam ser atendidas.*

Definição de interessados: Interessados: todos aqueles que tem algum tipo de interesse ou expectativa no projeto. Eles também podem ser reconhecidos como “clientes” internos ou externos, “audiências”, “parceiros” e outros.

3. Como foi a discussão sobre o envolvimento de interessados na fase de planejamento da avaliação?

4. O avaliador mencionou a importância disso? 03

Sim Não

5. Descreva como este aspecto foi tratado pelo avaliador.

6. Você acha importante o envolvimento de outros interessados visando ao atendimento de seus interesses? Dar exemplos: (Caso a caso.) 04

Sim Não

7. Cite um exemplo de interessados de seu Programa que não estiveram envolvidos na avaliação.

8. Você acha que teria sido importante o seu envolvimento? 05

Sim Não

9. Por que?

10. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

II - U5) Clareza dos relatórios: *Relatórios de avaliação devem descrever claramente o programa avaliado, incluindo seu contexto e os propósitos, procedimentos e conclusões da avaliação, de forma a prover informações essenciais que sejam facilmente entendidas.*

11. Houve algum acordo prévio sobre quantidade e formato dos relatórios?⁰⁶

Sim Não

Se Sim, quais foram?

12. Foi discutida a necessidade de desenvolver diferentes formatos para atender a diferentes públicos?⁰⁷

Sim Não

Se Sim, quais foram as sugestões e para que finalidade?

13. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

III- U6) Entrega a tempo e disseminação de relatórios: *Achados significativos e relatórios de avaliação devem ser disseminados junto aos usuários intencionais da avaliação, para serem utilizados no momento adequado.*

14. Como foram compartilhadas as informações referentes aos resultados parciais da avaliação durante o processo avaliativo? Com quem foram compartilhadas?

15. Como foi a relação do avaliador no que se refere aos prazos? Houve atraso?⁰⁸

Sim Não

Se Sim, os atrasos deveram-se a que?

16. As informações geradas pela avaliação foram disponibilizadas a tempo para alimentar processos decisórios?⁰⁹

Sim Não

Os eventuais atrasos trouxeram algum tipo de incômodo/prejuízo? Quais?

17. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

IV - U7- Impacto da avaliação: Avaliações devem ser planejadas, conduzidas e divulgadas de forma a estimular seu acompanhamento por parte dos principais interessados, aumentando assim as possibilidades de uso da avaliação.

18. O avaliador criou, desde o início da avaliação, estratégias para que os resultados da avaliação fossem utilizados pelos diferentes *interessados* do programa?¹⁰

Sim Não

Se Sim, quais foram as estratégias?

19. Quais foram os procedimentos de comunicação utilizados durante o processo avaliativo? Para quem?

20. Em que medida os usuários da avaliação ficaram satisfeitos com o impacto gerado pela avaliação? A avaliação foi útil?

21. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

V - V2 - Viabilidade política: *A avaliação deve ser planejada e conduzida tendo-se clareza antecipada dos diferentes posicionamentos dos vários grupos de interesse. Deve assim conseguir a cooperação de todos e tentar evitar, ou criar mecanismos para contra-atacar, possíveis esforços de qualquer destes grupos para dificultar a operacionalização, ou confundir, ou usar de forma inadequada os resultados da avaliação.*

22. Durante a fase de desenvolvimento da proposta avaliativa discutiu-se questões relacionadas a eventuais conflitos de interesse ou outras eventuais iniciativas que pudessem representar uma ameaça ao processo avaliativo ?¹¹

Sim Não

Se Sim, quais foram as questões levantadas?

23. Como foram tratadas as eventuais ações de forças políticas que interagiram durante o processo avaliativo?

24. As medidas adotadas para minimizar eventuais conflitos foram eficazes? ¹²

Sim Não

Se Não, na sua opinião, por quais motivos?

25. Você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

VI - V3) Custo-efetividade: *A avaliação deve ser eficiente e produzir informações de valor significativo, de forma que os recursos utilizados possam ser justificados.*

26. Como foi a relação do avaliador no que se refere aos recursos utilizados no processo avaliativo? Foram explicitados os prováveis benefícios com a avaliação? Houve preocupação em se maximizar os recursos despendidos?

27. As informações geradas como resultado do processo avaliativo justificaram os investimentos realizados em termos de relevância e utilidade (novas, geradoras de *insights*)?

28. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

VII - PR5- Avaliação completa e justa: *as avaliações devem ser completas e justas no momento de examinar e registrar os principais pontos positivos e limitações do programa sendo avaliado, possibilitando a valorização dos aspectos de sucesso do programa e correção das falhas existentes.*

29. A avaliação realizada pode ser considerada completa e justa? Abordou tanto pontos fortes quanto fracos do Programa ? 13

Sim Não

Se Não, quais aspectos deixaram de ser abordados e por que motivo (na sua opinião)?

30. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

VIII- PR6- Divulgação dos resultados: *os responsáveis pela avaliação devem assegurar que todas as pessoas e instituições afetadas pelo processo avaliativo, e aquelas com direito legal, tenham acesso aos resultados da avaliação em sua totalidade, assim como a possíveis limitações relevantes do processo.*

31. Quais interessados tiveram acesso aos resultados da avaliação?

32. Quem decidiu sobre o público a ser informado?

33. Quais foram os critérios utilizados para definir esse público?

34. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

VIII – PR8) Responsabilidade fiscal: a alocação e o desembolso dos recursos destinados à avaliação devem ser realizados de forma adequada, prudente e eticamente responsável pelo avaliador, para que se possa prestar contas dos gastos de maneira transparente.

35. Como foi a relação entre cliente e avaliador no que se refere ao uso dos recursos? Quais foram os procedimentos adotados?

36. Você ficou satisfeito com os procedimentos relacionados à prestação de contas? 14

Sim Não

Se Não, quais foram os pontos críticos relacionados a esses aspectos?

37. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

IX - PC2) Análise do contexto: o contexto no qual o programa está inserido deve ser descrito com o detalhamento necessário para que suas possíveis influências no programa possam ser identificadas.

38. Como foram tratadas as eventuais influências contextuais sobre os resultados do programa?

39. Foi proposto um mapeamento *a priori* de algum outro programa ou projeto que pudesse concorrer com as ações do programa? 15

Sim Não

40. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

X - PC4) Fontes de informação defensáveis: as fontes de informação usadas na avaliação devem estar descritas em detalhes, de forma a permitir a análise da adequação da informação coletada. 16

41. Houve discussões sobre eventuais vieses gerados pelas diferentes fontes de informação utilizadas na avaliação?

Sim Não

Se Sim, quais foram as discussões e os procedimentos para minimizar ou informar sobre esses vieses?

42. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

XI - PC9) Análise de informações qualitativas: informações qualitativas devem ser apropriada e sistematicamente analisadas de forma a possibilitar responder efetivamente às perguntas avaliativas formuladas.

43. Como foram tratadas as eventuais limitações das informações geradas pelo processo avaliativo no que diz respeito a limitações relacionadas aos instrumentos utilizados e análises realizadas?

44. Você acha que este *Standard* pode ser considerado como um parâmetro (dentro outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

***XII - PC12) Meta-avaliação:** o processo de desenvolvimento, bem como o possível impacto da avaliação para o programa, devem ser avaliados utilizando os Standards descritos aqui, ou outros, a fim de garantir que o processo transcorra de forma adequada e que, ao final, os principais interessados/afetados pela avaliação possam visualizar com clareza os pontos positivos e as limitações da avaliação.*

45. Na sua opinião, quais os motivos da não adoção de procedimentos de meta-avaliação?

46. Você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

ANEXO D - ROTEIRO DE ENTREVISTAS - AVALIADORES

Entrevistado : _____ Data: __/__/__ Cód. _____

Introdução

- Agradecimento pelo apoio a trabalho.
- Apresentação dos Objetivos da pesquisa.
- Descrição das estratégias da pesquisa: 1) 5 casos, 2) *Checklist*, 3) entrevista em profundidade com gerentes e pesquisadores, 4) grupo-foco externo.

Roteiro de Perguntas:

1. Você já conhecia os *Standards* para Avaliação de Programas antes da pesquisa?⁰¹

Sim

Não

Em caso positivo, foi feito uso dos *Standards* para a avaliação estudada nesta pesquisa?⁰²

Sim

Não

Em caso positivo, que tipo de uso foi feito?

Apresentação dos *Standards* :

- Foram criados entre 1989 e 1994 por 15 entidades: AEA, *American Educational Research Association*, pela *American Psychological Association* e pelo *National Council on Measurement in Education*.
- “Os *Standards* são princípios orientadores e não uma prescrição rígida. Eles contém orientações para evitar eventuais erros e fundamentam-se em práticas generalizadamente aceitas, além de propor diretrizes que refletem as **best-practices** em avaliação de programas atualmente”.
- “Os *Standards* devem ser utilizados como orientação para avaliar planos e relatórios... devem ser utilizados como meio para troca de informação sobre a qualidade da avaliação entre o cliente, os avaliadores e os vários stakeholders envolvidos...”

2. Reação imediata. Você acha que estes 30 *Standards* podem ser úteis para o aprimoramento da prática da avaliação de programas de investimento social privado? Em caso positivo, em que?
3. Apresentação do resultado do *Checklist*: A análise demonstra maiores pontuações nos seguintes *Standards*:.....
e menores nos seguintes:

I - U1) Identificação dos interessados: *Pessoas envolvidas ou afetadas pela avaliação devem ser identificadas, de forma que suas necessidades possam ser atendidas.*

Definição de interessados: Interessados: todos aqueles que têm algum tipo de interesse ou expectativa no projeto. Eles também podem ser reconhecidos como “clientes” internos ou externos, “audiências”, “parceiros” e outros.

Resultado do Checklist com relação a este Standard

U1 identificação dos interessados	NC	CP
O avaliador engajou líderes para identificar outros interessados	3	1
Os interessados ajudaram a identificar outros interessados	5	
Foram propostas estratégias para envolvimento dos interessados durante a avaliação no contrato	2	1

4. Com base na experiência do caso em estudo, comente sobre como foi tratado o envolvimento de interessados. Estimulou-se o envolvimento de outros interessados? Como a gerência recebeu a idéia de envolvimento de outros interessados?
5. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?
6. Tomando como base sua experiência na avaliação objeto deste estudo, bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

II - U5) Clareza dos relatórios: Relatórios de avaliação devem descrever claramente o programa avaliado, incluindo seu contexto e os propósitos, procedimentos e conclusões da avaliação, de forma a prover informações essenciais que sejam facilmente entendidas.

Resultado do Checklist com relação a este Standard

U5 Clareza dos relatórios	NC	CP
Foram empregadas diferentes mídias para informar diferentes audiências	1	2
Foram apresentados exemplos para facilitar a compreensão das audiências	2	1

7. Com base na experiência do caso em estudo, comente sobre como foram estruturados os relatórios quanto a formatos para comunicação dos resultados da avaliação?
8. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?
9. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

III - U6) Entrega a tempo e disseminação de relatórios: Achados significativos e relatórios de avaliação devem ser disseminados junto aos usuários intencionais da avaliação, para serem utilizados no momento adequado.

Resultado do Checklist com relação a este Standard

U6 Entrega a tempo e disseminação de relatórios	NC	CP
O avaliador procurou, em conjunto com o cliente, identificar, acessar e informar todos os usuários intencionais sobre os achados da avaliação	1	2
O relatório final foi entregue no prazo previsto	2	1
O avaliador desenvolveu <i>releases</i> para divulgação dos achados na mídia	4	1

10. Com base na experiência do caso em estudo, comente sobre como foram os fluxos de informações parciais sobre os achados da avaliação durante o processo avaliativo.
11. Comente sobre as eventuais dificuldades com prazos. A que se deveram? Acarretaram incômodos/prejuízos?
12. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?
13. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

IV - U7) Impacto da avaliação: *Avaliações devem ser planejadas, conduzidas e divulgadas de forma a estimular seu acompanhamento por parte dos principais interessados, aumentando assim as possibilidades de uso da avaliação.*

Resultado do Checklist com relação a este Standard

U7 Impacto da avaliação	NC	CP
O avaliador criou sistemática para prover <i>follow-up</i> e suporte para interpretação e aplicação dos achados	1	2
O avaliador demonstrou de forma clara como os achados da avaliação poderiam ser incorporados no trabalho da equipe	1	2
O avaliador suplementou relatórios escritos com constantes comunicações verbais	1	2
O avaliador conduziu reuniões de <i>feedback</i> para ir além e viabilizar a aplicação dos achados	1	2

14. Com base na experiência do caso em estudo, comente sobre as estratégias utilizadas para estimular o uso dos achados da avaliação pelos diferentes interessados.
15. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?

16. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

V - V2) Viabilidade política: *A avaliação deve ser planejada e conduzida tendo-se clareza antecipada dos diferentes posicionamentos dos vários grupos de interesse. Deve assim conseguir a cooperação de todos e tentar evitar, ou criar mecanismos para contra-atacar, possíveis esforços de qualquer destes grupos, para dificultar a operacionalização, ou confundir, ou usar de forma inadequada os resultados da avaliação.*

Resultado do Checklist com relação a este Standard

V2 Viabilidade política	NC	CP
O avaliador antecipou-se às diferenças de posições de diferentes grupos	0	3
O avaliador reportou divergências de pontos de vista	0	3

17. Com base na experiência do caso em estudo, comente sobre como foram tratadas as questões relacionadas a eventuais conflitos de interesse ou outras eventuais iniciativas que pudessem representar uma ameaça ao processo avaliativo.
18. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?
19. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

VI - V3) Custo-efetividade: A avaliação deve ser eficiente e produzir informações de valor significativo, de forma que os recursos utilizados possam ser justificados.

Resultado do Checklist com relação a este Standard

V3 Custo-efetividade	NC	CP
O processo avaliativo produziu informações novas, úteis à tomada de decisão	0	3
O processo avaliativo fomentou melhorias no programa	0	3
O processo avaliativo produziu informações sobre o <i>accountability</i> (uso responsável de recursos e prestação de contas às autoridades superiores ligadas ao programa. Exemplo: financiadores e Conselho Diretor)	0	3
O processo avaliativo propiciou a geração de novos <i>insights</i> sobre o programa	1	3

20. Comente sobre sua percepção sobre a relação custo-efetividade da avaliação objeto desta pesquisa.

21. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?

22. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

VII - PR5) Avaliação completa e justa: as avaliações devem ser completas e justas no momento de examinar e registrar os principais pontos positivos e limitações do programa avaliado, possibilitando a valorização dos aspectos de sucesso do programa e a correção das falhas existentes.

Resultado do Checklist com relação a este Standard

P5 Avaliação completa e justa	NC	CP
O avaliador reconheceu as limitações do relatório final	4	1
O avaliador apresentou no relatório as limitações do processo avaliativo para o julgamento de determinados aspectos do programa	4	1

23. Comente sobre a abrangência da avaliação no que diz respeito ao levantamento dos pontos fortes e fracos do Programa.

24. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?

25. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

VIII - PR6) Divulgação de resultados: os responsáveis pela avaliação devem assegurar que todas as pessoas e instituições afetadas pelo processo avaliativo e aquelas com direito legal tenham acesso aos resultados da avaliação em sua totalidade, assim como a possíveis limitações relevantes do processo.

Resultado do Checklist com relação a este Standard

PR6 Divulgação dos achados	NC	CP
O avaliador assegurou que os relatórios chegassem a diferentes audiências	0	4

26. Comente sobre o acesso e divulgação dos resultados da avaliação junto aos diferentes interessados. Quais interessados tiveram acesso aos resultados da avaliação?
27. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?
28. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

VIII - PR8) Responsabilidade fiscal: a alocação e o desembolso dos recursos destinados à avaliação devem ser realizados de forma adequada, prudente e eticamente responsável pelo avaliador, para que se possa prestar contas dos gastos de maneira transparente.

Resultado do *Checklist* com relação a este *Standard*

PR8 Responsabilidade fiscal	NC	CP
Foi incluído no relatório final um informe resumido sobre as despesas realizadas	5	0

29. Comente sobre a relação entre avaliador e cliente e em relação ao uso dos recursos. Quais foram os procedimentos adotados? Houve prestação de contas? O cliente ficou satisfeito com os procedimentos adotados?
30. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?
31. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

IX - PC2) Análise do contexto: o contexto no qual o programa está inserido deve ser descrito com o detalhamento necessário para que suas possíveis influências no programa possam ser identificadas.

Resultado do Checklist com relação a este Standard

PC2 Análise do contexto	NC	CP
Foram identificados e descritos quaisquer competidores (projeto, programa) que ocorreram no mesmo período e ambiente do programa	1	2

32. Comente sobre o tratamento metodológico dado às eventuais influências contextuais sobre os resultados do programa.
33. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?
34. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

X - PC4) Fontes de informações defensáveis: as fontes de informação usadas na avaliação devem ser descritas em detalhes, de forma a permitir a análise da adequação da informação coletada.

Resultado do Checklist com relação a este Standard

PC4 Fontes de informações defensáveis	NC	CP
O avaliador registrou e reportou ocorrências que possam ter sido geradoras de vieses	3	1

35. Comente sobre o tratamento metodológico dado à seleção e descrição sobre as diferentes fontes de informação utilizadas na avaliação.
36. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?

37. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

XI - PC6) Informações confiáveis: *os procedimentos para coleta de informações devem ser escolhidos ou desenvolvidos e depois aplicados de forma a assegurar que a informação obtida seja suficientemente confiável.*

Resultado do Checklist com relação a este Standard

PC6 Informações confiáveis	NC	CP
O avaliador reportou fatores que o influenciaram (características de análise, condições em que foi feita a coleta de dados e vieses dos avaliadores)	2	1

38. Comente sobre o tratamento metodológico dado à seleção e descrição sobre as diferentes fontes de informação utilizadas na avaliação.
39. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?
40. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

XII - PC9) Análise de informações qualitativas: informações qualitativas devem ser apropriada e sistematicamente analisadas de forma a possibilitar responder efetivamente às perguntas avaliativas formuladas.

Resultado do Checklist com relação a este Standard

PC9 Análise de informações qualitativas	NC	CP
O avaliador reportou limitações sobre as fontes de informações, análise e inferências	3	1

41. Comente sobre o tratamento dado às eventuais limitações dos instrumentos e análises realizados?
42. Comente sobre os resultados da tabela. O que os resultados do *Checklist* nos mostram?
43. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist*, você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

XIII - PC12) Meta-avaliação: o processo de desenvolvimento, bem como o possível impacto da avaliação para o programa, devem ser avaliados utilizando os Standards descritos aqui, ou outros, a fim de garantir que o processo transcorra de forma adequada e que, ao final, os principais interessados/afetados pela avaliação possam visualizar com clareza os pontos positivos e as limitações da avaliação.

44. Na sua opinião, quais os motivos da não adoção de procedimentos de meta-avaliação em quatro dos cinco casos?
45. Tomando como base sua experiência na avaliação objeto deste estudo bem como a análise do resultado do *Checklist* você acha que este *Standard* pode ser considerado como um parâmetro (dentre outros) justo de julgamento de uma avaliação de programa na realidade brasileira?

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso de discordância.

ANEXO E - QUESTIONÁRIOS DE CONCORDÂNCIA

Nome :

Programa:

Cod.

Questionário

Pergunta : Os 30 *Standards* definidos pelo *Joint Committee of Standards for Educational Evaluation* dos E.U.A podem ser aceitos como diretrizes para orientar práticas avaliativas bem como avaliar a qualidade de avaliações de projetos e programas de investimento social privado no Brasil?

Escolha uma alternativa que manifeste sua concordância ou discordância com relação a cada um dos *Standards*.

1. Atributo Utilidade

U2) Credibilidade do avaliador: As pessoas que conduzem a avaliação devem ser confiáveis e competentes, a fim de que os resultados da avaliação obtenham grau elevado de credibilidade e aceitação.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

U3) Escopo e seleção de informações: Informações coletadas devem ser selecionadas de forma ampla para que possam abordar perguntas pertinentes sobre o programa e trazer respostas a necessidades e interesses dos clientes e de outros interessados pela avaliação.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

U4) Identificação de valores: As perspectivas, os procedimentos e as justificativas usados para interpretar os resultados da avaliação devem ser cuidadosamente descritos, de forma que as bases para julgamentos de valores fiquem bem claras.

Concordo			Discordo		
Pouco	Muito	Totalmente	Muito	Pouco	Totalmente

Justifique em caso se discordância:

2. Atributo Viabilidade

V1) Procedimentos práticos: Os procedimentos da avaliação devem ser práticos a fim de evitar ao máximo perturbações no momento em que as informações necessárias estão sendo coletadas.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

3 - Atributo Propriedade

PR1) Orientação para o serviço:

avaliações devem ser desenhadas de forma a ajudar as organizações a atender e responder de forma eficaz às necessidades de todas as pessoas e instituições servidas ou afetadas de alguma maneira por seus programas.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PR2) Acordos formais:

acordos firmados sobre “o quê”, “como”, “quem” e “quando” fazer as atividades pertinentes à avaliação devem estar registrados num documento, de modo que as partes envolvidas sintam-se na obrigação de honrar os compromissos assumidos ou renegociá-los formalmente se for necessário.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PR3) Direitos dos indivíduos:

avaliações devem ser planejadas e levadas a cabo considerando sempre o respeito e o zelo pelos direitos individuais das pessoas nelas envolvidas ou afetadas por seus resultados.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PR4) Relações humanas: os avaliadores, no momento de se relacionar com as outras pessoas envolvidas na avaliação, devem respeitar a dignidade e o valor dessas pessoas, para que elas não se sintam em nenhum momento ameaçadas ou prejudicadas.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PR7) Conflito de interesses: qualquer conflito de interesses durante o desenvolvimento do processo avaliativo deve ser tratado de forma aberta e honesta entre as partes envolvidas, de forma a não comprometê-lo ou a seus resultados.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

4. Atributo Precisão

PC1) Documentação do programa: o programa avaliado deve ser descrito e documentado de forma clara e precisa, para que possa ser facilmente identificado.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PC3) Descrição de propósitos e procedimentos: os propósitos e procedimentos da avaliação devem ser descritos de maneira clara e monitorados para que sejam facilmente identificados e examinados.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PC5) Informações válidas: os procedimentos para a coleta de informações devem ser escolhidos ou desenvolvidos e depois aplicados de forma a assegurar que a interpretação de resultados seja válida e útil.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PC6) Informações confiáveis: os procedimentos para coleta de informações devem ser escolhidos ou desenvolvidos e depois aplicados de forma a assegurar que a informação obtida seja suficientemente confiável.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PC7) Informação sistemática: A informação coletada, processada e escrita em relatórios deve ser sistematicamente revisada e qualquer erro encontrado deve ser corrigido.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PC8) Análise de informações quantitativas: informações quantitativas deverão ser apropriada e sistematicamente analisadas de forma a possibilitar responder efetivamente às perguntas avaliativas formuladas.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PC10) Conclusões justificáveis: as conclusões de uma avaliação devem ser explicitamente justificadas para que possam ser analisadas pelos principais interessados/afetados pela avaliação e/ou pelo programa.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

PC11) Imparcialidade de relatórios: os procedimentos para elaboração de relatórios deverão incluir métodos que previnam possíveis distorções causadas por sentimentos/opiniões/vieses pessoais, a fim de que os relatórios reflitam de forma justa os resultados da avaliação.

Concordo			Discordo		
Pouco	Muito	Totalmente	Pouco	Muito	Totalmente

Justifique em caso se discordância:

8 REFERÊNCIAS BIBLIOGRÁFICAS

AUSTIN, J., E. **Parcerias- Fundamentos e benefícios para o terceiro setor**. São Paulo, Futura, 2001.

CAMPBELL, D., T. **Degrees of freedom and the Case Study**. Comparative Political Studies, V 8,178-193

CHIANCA, T.; MARINO, E; SCHIESARI, L. **Desenvolvendo a cultura de avaliação em organizações da sociedade civil**. São Paulo, Editora Global e Instituto Fonte, 2000.

CRONBACH, L. J. & SUPPES, P. **Research for tomorrow's Schools: Disciplined Inquiry for Education**. New York: Macmillan, 1969

Evaluation Research Society. **Standards for Evaluation** (draft). Washington, DC:ERS, 1980.

FALCONER, A. P. **A promessa do terceiro setor – Um estudo sobre a Construção do Papel das Organizações Sem fins lucrativos e do seu Campo de gestão**. São Paulo, Dissertação (Mestrado) – Faculdade de Economia e Administração e Contabilidade. Universidade de São Paulo, 1999. 153 p.

FALCONER, A. P. **Um setor ou diversos? Reconhecendo o 3º setor no BR**. III SEMEAD.

FERNANDES, R. C. O que é o Terceiro Setor? in **3º Setor – Desenvolvimento Social Sustentado**. Rio de Janeiro, Paz e Terra, 1994.

FETTERMAN, M. D. **Empowerment Evaluation**. Newbury Park, CA: SAGE Publications, Inc.

FISCHER, R., M. **O desafio da colaboração: práticas de responsabilidade entre empresas e terceiro setor**. São Paulo, Editora Gente, 2002.

FREITAG, B. & ROUNET, P. S. ORGS. **Habermas, Sociologia**. São Paulo, Editora Ática, 2001.

GIFE - Grupo de Institutos, Fundações e Empresas. **Investimento Social Privado no Brasil**. São Paulo, Editora Fundação Peirópolis, 2001.

- GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo, Editora Atlas, 2002.
- GUBA, E. G., & LINCOLN, Y. S. **Effective Evaluation**. San Francisco: Jossey Bass, 1981.
- GUBA, E. G., & LINCOLN, Y. S. **Fourth Generation Evaluation**. Newbury Park, CA: SAGE Publications, Inc., 1989.
- HOUSE, E. R. **Evaluation with validity**. Beverly Hills, CA: SAGE, 1980
- HOUSE, E. R. **Philosophy of Evaluation**. New Directions for Program Evaluation, San Francisco, Jossey Bass,. 1983. V19.
- HOUSE, R., E & HOWE, R. K. **Deliberative Democratic Evaluation**. **New Directions for Evaluation**. São Francisco, Jossey Bass Publishers, 2000. V. 85, p 3-11.
- INSTITUTO DE PESQUISAS ECONÔMICAS. **A Iniciativa Privada e o Espírito Público - Resultados Nacionais**. Brasília. Disponível em < www.ipea.gov.br>. Acesso em 10/10/2002.
- JANG, S. **The appropriateness of Joint Committee Standards in Non-Western Settings: A Case of South Korea**. Syracuse University, 2000.
- JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION **The program evaluation Standards**, 2.ed., Thousand Oaks: Sage, 1994.
- KUHN, T. **A Estrutura das Revoluções Científicas**. São Paulo, Editora Perspectiva, 2000.
- LANDIM, L.M. **Defining Nonprofit Sector in Brazil**, The Johns Hopkins University, 1993.
- LANDIM, L.M. **The Nonprofit Sector in Brazil, in The Nonprofit Sector in Developing World**. Anheier & Salamon, L. (Eds.) Manchester/New York, Manchester University Press, 1998.
- LOGAN, D. et al. **Global Corporate Citizenship – Rationale and Strategies**. Washington D. C., The Hitachi Foundation, 1997.
- MARINO, E. **How Can Evaluation and Corporate Volunteerism Add Value to Corporate Citizenship in Brazil?**.(Working paper – não publicado) New York, City University of New York, 1998.
- MORAES, M., C. **O paradigma educacional emergente**. Campinas, Papirus, 1997.
- MORIN, E. **Epistemologia da Complexidade**. *In*. D.F. Schnitman (org.). **Novos Paradigmas, Cultura e Subjetividade**. Porto Alegre: Artes Médicas, 1996.

NOWAKOWSKI, R., J. **An Interview with Ralph Tyler**. Ocasional Paper Series. Western Michigan University, 1981.

PATTON, M. Q. **Practical Evaluation**. Newbury Park, CA, Sage, 1982

PATTON, M.Q. **Qualitative research and evaluation methods**. 2.ed., Newbury Park, CA, Sage, 1990.

PATTON, M.Q. **Utilization Focused-Evaluation**. Newbury Park, CA, Sage, 1996

PATTON, M.Q. **Qualitative research and evaluation methods**. 3.ed., Newbury Park, CA: Sage, 2002.

PELIANO, A. M. T. M. org. **Um retrato da ação social das empresas do Sudeste brasileiro**. IPEA, Brasília, 2000

PELIANO, A. M. T. M. org. **A iniciativa Privada de Espírito Público**. IPEA, Brasília, 2000.

PRESKILL, H & TORRES, T. R, The Learning Dimension of Evaluation Use. **New Directions for Evaluation**. San Francisco, Jossey Bass, 2000. V. 88, p. 25 - 37.

RENZ, L. et al. **Yearbook of facts and figures on private, corporate and community foundations**. New York, Foundation Center, 1997.

RIDDELL, R. **Linking Costs and Benefits in NGO Development Projects: A study by the Overseas Development Institute commissioned by the Development Administration**. ODI, London, 1997.

ROCHE Chris. **Avaliação de Impacto dos Trabalhos de ONGs: aprendendo a valorizar as mudanças**. São Paulo, Cortez Editora, 2000.

RUCKLE, J. **Distinctive qualities of third sector organizations**. New York and London, Garland Publishing, 1993.

RUSSOM, C. **The Program Evaluation Standards in International Settings**. The Evaluation Center Occasional Papers. Kalamazoo, 2001.

SALAMON, L. M. & ANHEIER, H. **America's Nonprofit Sector- A Primer**. New York, Foundation Center, 1992.

SALAMON, L. M. & ANHEIER, H. **The Emerging Sector**. The Johns Hopkins University, 1994.

SALAMON, L. M. & ANHEIER, H. **The nonprofit Sector in developing world**. The Johns Hopkins University, 1998.

- SANDERS J. R. **A basis for determining the adequacy of evaluation designs.** Occasional Paper Series. Alaska Department of Education to the Northwest Regional, 1976.
- SCHRAMM, W. **Notes on Case Studies of Instructional Media Projects.** Working Paper, the Academy of Educational Development, Washington, DC, 1971.
- SCRIVEN, M. **An introduction to Meta-evaluation.** Educational Product Report. V. 2, N. 5, 1969
- SCRIVEN, M. **Evaluation Thesaurus.** Sage, Newbury Park, 1991.
- SCRIVEN, M. **Goal Free Evaluation. Evaluation News & Comment.** Australian Evaluation Society, 1996. V 5, Number 2, p. 12.
- SELLTIZ, et alii. **Métodos de Pesquisa nas Relações Sociais.** São Paulo, Pedagógica e Universitária, 1985.
- SINGER D., B. **Towards a Sociology of Standards: Problems of a Criterial Society.** The Canadian Journal of Sociology, 1996. V 21, n 2, p. 203-221.
- STAKE R. E. **The countenance of Educational Evaluation.** Teachers College Record, 68, 523-540, 1967.
- STAKE R., E. **Program Evaluation Particularly Responsive.** Center for Instructional Research and Curriculum Evaluation. University of Illinois at Urbana-Campaign, 1975.
- STAKE R. E.. **Summary of evaluation of reader focused writing for veterans benefits administration.** American Journal of evaluation, V. 20, 323-343, 1999.
- STEVENSON, W.,J. **Estatística aplicada a administração.** São Paulo, HARBRA, 1981.
- STEWART, D. W.; SHAMDASANI, P. N. **Focus group, theory and practice.** Newbury, California. Sage Publications, Inc, 1990.
- STUFFLEBEAM, D. L. **Meta-evaluation.** Occasional Papers series. The Evaluation Center, Western Michigan University, Kalamazoo, 1974.
- STUFFLEBEAM, D. L. **The CIPP Model for Evaluation.** Texto não publicado. MI: The Evaluation Center, Western Michigan University, Kalamazoo, 1999.
- STUFFLEBEAM, D. L. **Program Evaluation Metaevaluation Checklist** – based on The Program Evaluation Standards. (pdf version). The Evaluation Center. Western Michigan University. Kalamazoo, 2000.

STUFFLEBEAM, D. L. **Evaluation Models. New Directions for Evaluation**, San Francisco, Jossey Bass,. 2001a. V. 89, p. 7 – 98.

STUFFLEBEAM, D. L. **The Metaevaluation Imperative. The American Journal of Evaluation**. New York, Elsevier Science Inc., 2001b. V. 22, p. 183 – 209.

STUFFLEBEAM, D., MADAUS, G., & KELLAGHAN, T. (Eds.). **Evaluation Models. The CIPP Model for Evaluation by Daniel L. Stufflebeam**, (tradução Chianca, T., K. no prelo) Kluwer Academic Publishers - Boston / Dordrecht / Londres, 2000. p. 279-317

TAUT, S. **Cross- cultural transferability of the program evaluation Standards**. The Evaluation Center. Western Michigan University, 2001.

THIOLLENT, M. **Metodologia da Pesquisa - ação**. 11. ed. São Paulo, Cortez Editora, 2000.

WEEDEN, CURT.. **Corporate Social Investing: the breakthrough strategy for giving and getting corporate contribution**. São Francisco, CA. Berret-Koehler Publishers, Inc., 1998.

WIDMER, T. **Evaluating evaluations: does the Swiss practice live up to the “program evaluation Standards”?** Institute of Political Science. University of Zurich, 1995.

WORTHEN, SANDERS & FITZPATRICK. **Educational Evaluation – Alternative Approaches and Practical Guidelines**. 2. Ed., USA, Longman Publishers, 1987.

WORTHEN, SANDERS & FITZPATRICK. **Program Evaluation: Alternative approaches and practical guidelines**. 2. Ed., USA, Longman Publishers, 1997

YIN, R. K. **Estudo de caso – planejamento e métodos**. Porto Alegre, Bookman, 2002.