

## Avaliação de Programa: aspectos teóricos e metodológicos no contexto da agricultura familiar



ISSN 1517-5111

Abril, 2008

*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Cerrados  
Ministério da Agricultura, Pecuária e Abastecimento*

## **Documentos 211**

# **Avaliação de Programa: aspectos teóricos e metodológicos no contexto da agricultura familiar**

*Francisco Eduardo de Castro Rocha  
Francisco José Batista de Albuquerque  
Maria Quitéria dos Santos Marcelino*

Embrapa Cerrados  
Planaltina, DF  
2008

Exemplares desta publicação podem ser adquiridos na:

**Embrapa Cerrados**

BR 020, Km 18, Rod. Brasília/Fortaleza

Caixa Postal 08223

CEP 73310-970 Planaltina, DF

Fone: (61) 3388-9898

Fax: (61) 3388-9879

<http://www.cpac.embrapa.br>

[sac@cpac.embrapa.br](mailto:sac@cpac.embrapa.br)

**Comitê de Publicações da Unidade**

Presidente: *José de Ribamar N. dos Anjos*

Secretária-Executiva: *Maria Edilva Nogueira*

Supervisão editorial: *Fernanda Vidigal Cabral de Miranda*

*Francisca Elijani do Nascimento*

*Jussara Flores de Oliveira Arbués*

Revisão de texto: *Fernanda Vidigal Cabral de Miranda*

Normalização bibliográfica: *Rosângela Lacerda de Castro*

Tratamento de ilustrações: *Wellington Cavalcanti*

Editoração eletrônica: *Wellington Cavalcanti*

Capa: *Wellington Cavalcanti*

Foto(s) da capa: *Francisco Eduardo de Castro Rocha*

Impressão e acabamento: *Divino Batista de Sousa*

*Alexandre Moreira Veloso*

Impresso no Serviço Gráfico da Embrapa Cerrados

**1ª edição**

1ª impressão (2008): tiragem 100 exemplares

**Todos os direitos reservados**

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP)  
Embrapa Cerrados**

---

R672a Rocha, Francisco Eduardo de Castro.

Avaliação de programa: aspectos teóricos e metodológicos no contexto da agricultura familiar / Francisco Eduardo de Castro Rocha, Francisco José Batista de Albuquerque, Maria Quitéria dos Santos Marcelino. – Planaltina, DF : Embrapa Cerrados, 2008.

62 p. – (Documentos / Embrapa Cerrados, ISSN 1517-5111 ; 211)

1. Psicologia social. 2. Programa social. 3. Pronaf. I. Albuquerque, Francisco José Batista de. II. Marcelino, Maria Quitéria dos Santos. III. Título. IV. Série.

302 - CDD 21

# **Autores**

**Francisco Eduardo de Castro Rocha**

Engenheiro Agrícola, Psicólogo, D.Sc.

Pesquisador da Embrapa Cerrados

rocha@cpac.embrapa.br

**Francisco José Batista de Albuquerque**

Psicólogo, D.Sc.

Professor da Universidade Federal da Paraíba

frajoba@uol.com.br

**Maria Quitéria dos Santos Marcelino**

Pedagoga, D.Sc.

Universidade Federal da Paraíba

mqmarcellino@yahoo.com.br

# Agradecimento

Agradeço à Embrapa pelo apoio; à Universidade Federal da Paraíba, assim como ao Núcleo de Estudos sobre Desenvolvimento Rural e Avaliação de Programas Sociais pelo acolhimento; à Secretaria Executiva Estadual do Pronaf da Paraíba; à Emater-PB e ao Banco do Nordeste do Brasil pela atenção recebida por parte dessas instituições.

# Apresentação

A pesquisa agropecuária na Embrapa ainda tem sido predominantemente de cunho tecnológico. No entanto, na Embrapa Cerrados, há tempos são conduzidos projetos de política local, com foco no desenvolvimento rural sustentável, como o Projeto Silvânia (1987 - 1998). Esse trabalho foi conduzido com pequenos produtores e assentados no Município de Silvânia, GO, tendo como principal resultado a criação de uma central de associações, a qual, posteriormente, serviu de base para a formação de uma cooperativa, que hoje responde por quase toda a economia local.

Nesse contexto, não tem sido observada de forma sistemática a avaliação de projetos e, menos ainda, a avaliação de políticas públicas que envolvem grande contingente de beneficiários. Portanto, considera-se que as informações teóricas e metodológicas disponibilizadas nesta publicação vão ao encontro das necessidades atuais da Embrapa, que busca sempre melhorar seus processos e produtos de trabalho. Nesse sentido, são bem-vindas as novas possibilidades de pesquisas interativas com as distintas áreas do conhecimento e complementares às agrárias.

Pretende-se, assim, com este trabalho, lançar novas idéias e informações de tal forma que possam somar-se aos trabalhos da

Embrapa, tanto para a coordenação quanto para a participação em projetos sociais, utilizando, para isso, estudos científicos realizados no âmbito da avaliação de programa. A aplicação desse tipo de conhecimento é de relevância não somente para trabalhos de avaliação de projetos de cunho social da empresa ou de políticas públicas, mas também para pesquisas voltadas à área de transferência de tecnologia, campo este que necessita ser avaliado continuamente e compreendido por esse prisma.

*Roberto Teixeira Alves*  
Chefe-Geral da Embrapa Cerrados

# Sumário

Introdução.....	11
Avaliação de programa: uma estratégia de correção de rumos .....	14
Conceitos básicos: um olhar contextual .....	16
Aspectos teóricos e metodológicos: uma breve discussão .....	29
Avaliação de planos, programas e projetos: uma diversidade de métodos e modelos .....	36
Referências .....	56
Referências recomendadas .....	61
Abstract .....	62

# Avaliação de programa: aspectos teóricos e metodológicos no contexto da agricultura familiar<sup>1</sup>

---

*Francisco Eduardo de Castro Rocha*

*Francisco José Batista de Albuquerque*

*Maria Quitéria dos Santos Marcelino*

## Introdução

Nesta publicação, pretende-se mostrar alguns aspectos teóricos e metodológicos que serviram de base para a avaliação do Programa Nacional de Fortalecimento da Agricultura Familiar (Pronaf) na Paraíba. A necessidade de os gestores obterem informações, por meios mais confiáveis e detalhados, para embasarem suas intervenções, preferencialmente a um menor custo de operação, deve passar por um processo mais detalhado de investigação como, por exemplo, a avaliação de programa.

O governo federal lançou o Pronaf em 1996 para apoiar financeiramente, por meio da concessão do crédito rural, atividades agropecuárias e não-agropecuárias exploradas mediante emprego direto da força de trabalho do produtor rural e de sua família. As atividades não-agropecuárias referem-se aos serviços vinculados ao turismo rural, à produção artesanal, ao agronegócio familiar e a outras prestações de serviços no meio rural, compatíveis com a natureza da exploração rural e com o melhor emprego da mão-de-obra familiar ([BANCO CENTRAL DO BRASIL, 2007](#)). Como conseqüência, espera-se a promoção do

---

<sup>1</sup> Parte da tese do primeiro autor - (Avaliação psicossocial do Programa Nacional de Fortalecimento da Agricultura Familiar (Pronaf) no Estado da Paraíba, 2008)

desenvolvimento rural sustentável e o fortalecimento da agricultura familiar por meio de seus beneficiários: os agricultores familiares, assentados, povos indígenas e pescadores, estruturados em cinco grupos de beneficiários: A, B, A/C, D e E ([AMORIM, 2003](#); [FERREIRA, 2007a,b](#)).

Dados do [IBGE \(1995\)](#) permitem verificar que existem 4,3 milhões (89,3 %) de estabelecimentos agrícolas de 0 ha a 100 ha (agricultura familiar) vs. 516 mil (10,7 %) estabelecimentos maiores que 100 ha (agricultura patronal). No contexto do Pronaf, segundo Brasil (2007a), o governo federal, para a safra 2006/2007, liberou R\$ 8,4 bilhões para atender a 1,7 milhão de contratos. Para a safra 2007/2008, disponibilizou a cifra de R\$ 12 bilhões, na forma de dotação orçamentária. Por meio desses dados, pode-se inferir que o Pronaf (1,7 milhão de contratos) ainda não conseguiu atender sequer a metade da população potencial de agricultores familiares (4,3 milhões), o que reforça sua importância perante esse segmento da agricultura, mais do ponto de vista comportamental que tecnológico.

Segundo [Brasil \(2005\)](#), no que diz respeito ao pagamento do Pronaf, 300 mil contratos de crédito rural foram renegociados em 2004 pelo governo federal. Em 2003, outras 500 mil operações também foram contempladas, totalizando cerca de R\$ 3,3 bilhões repactuados com os agricultores em todo o País. Como forma de minimizar esse problema, a resolução do Conselho Monetário Nacional (CMN), em 2007, definiu regras para a individualização e a renegociação das dívidas dos beneficiários do Pronaf e do antigo Programa Especial de Crédito para a Reforma Agrária (Procera) – que acaba de ser reformulada ([BRASIL, 2007b](#)).

[Volpi e Bressan \(2001\)](#), [Berger \(1973\)](#), [Albuquerque e Mascareño \(1999\)](#) consideram que problemas como carência de informações sobre o programa, idade avançada, baixo nível de escolaridade, pouca experiência em lidar com operações bancárias – por exemplo, o crédito rural –, cultivo de subsistência, baixa condição de manutenção da família, baixo nível de organização, falta de assistência técnica, época de liberação dos créditos e fatores edafoclimáticos devem ser

superados para que os agricultores possam honrar seus compromissos e alcançar um nível de vida compatível com o esforço dispendido para produzir alimentos para si e para o mercado.

A gestão dessa política, dentro de um campo de forças, é feita em diversos níveis institucionais. Se, por um lado, encontram-se os beneficiários habituados a receber benefícios a fundo perdidos, por outro, está o governo operacionalizando um programa de negócio, com previsão de retorno, tanto financeiro como social, para esses mesmos indivíduos.

Para planejar e organizar ações com maior probabilidade de êxito, considerando essa situação de complexidade, torna-se necessário um trabalho especializado e cuidadoso de investigação, com base no levantamento e na medição de informações provenientes dos envolvidos no programa. Essa é uma estratégia, entre várias outras, de se organizar para enfrentar problemas dessa natureza, levando-se em conta o feedback dos beneficiários, em termos de conhecimento, visão, experiência, criatividade (Fig. 1). Nesse caso, fica claro que não é uma ação do gestor para avaliar os beneficiários, mas é o beneficiário contribuindo de forma participativa na avaliação da política.

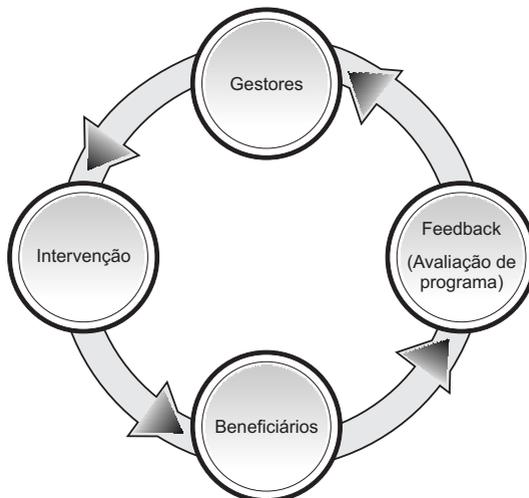


Fig. 1. Representação gráfica do ciclo de operacionalização do Pronaf.

Diante desse problema, optou-se, nesta publicação, por apresentar um estudo teórico sobre avaliação de programa, já que também dá suporte a outros trabalhos de pesquisa tanto no campo da avaliação de programas sociais quanto no de projetos, os quais podem ser considerados a base de funcionamento da Embrapa.

No ambiente rural, serve como exemplo de avaliação de resultado de programa, uma agroindústria para o processamento de frutas, no Assentamento São João II, em Pombal, conforme fotografia apresentada na capa da presente publicação. Pelo menos, até a data da coleta de dados (17 de fevereiro de 2006), ela se encontrava construída e equipada, mas sem previsão de iniciar seu funcionamento.

## **Avaliação de programa: uma estratégia de correção de rumos**

O Estado utiliza diversas estratégias para implementar as políticas públicas, por exemplo, os programas sociais. No caso das políticas agrícolas, o resultado esperado desses programas é o desenvolvimento rural obtido por meio de metas produtivistas e mecanismos como o crédito agrícola. Mas, subjacentes a esse processo, encontram-se as questões comportamentais, que são imprescindíveis ao andamento das atividades, pois o comportamento requerido afeta diretamente a eficiência e a eficácia com que são realizadas as ações. Para tanto, a avaliação torna-se um dos meios facilitadores do processo de intervenção do Estado.

Pouco se conhece da avaliação do Pronaf, principalmente, do ponto de vista comportamental. [Cozby \(2003\)](#) enfatiza que pesquisadores freqüentemente se defrontam com questões de pesquisa aplicada e de avaliação do comportamento. Nesse contexto, encontram-se as pesquisas de avaliação de programas voltadas para as reformas sociais e inovações introduzidas no governo, na educação, no sistema judiciário, na indústria, na assistência à saúde etc. Muitas pesquisas aplicadas são realizadas por grandes empresas, companhias de pesquisa de mercado, agências governamentais e agências que realizam pesquisas de opinião pública, e seus resultados não são publicados.

Independentemente de eles serem ou não publicados, são usados para ajudar pessoas a tomar decisões melhores em relação a problemas que requerem ação imediata.

Nessa perspectiva, valem como exemplo os atuais trabalhos do Tribunal de Contas da União (TCU), instituição governamental que sempre esteve ligada à auditoria dos recursos públicos, mas agora tem ampliado sua atuação ao desempenhar trabalhos na linha de avaliação de programa, considerando tanto os modelos lógicos como os de causa e efeito, delineamentos mais complexos ([BRASIL, 2000, 2001, 2002](#)).

Avaliar programas e políticas públicas constitui um poderoso instrumento para gestores, formuladores e implementadores, uma vez que serve de base para tomar decisões com maior objetividade e qualidade, maximizar os efeitos do gasto público, identificar e superar pontos de estrangulamento, verificar os êxitos dos programas e, conseqüentemente, abrir perspectivas racionais à implementação de políticas com maior capacidade para alcançar os resultados desejados. Dessa forma, o objetivo primordial da avaliação de um programa é contribuir para o seu aperfeiçoamento, por meio de informações indispensáveis para os gestores públicos, na qualidade de tomadores de decisão (*policy makers*), e aumentar o grau de satisfação de seus beneficiários ou usuários ([SILVA, 1999](#)).

Para isso, esse autor recomenda conhecer o programa, apreciar seu delineamento, seu modo de implementação na teoria e na prática. Conhecer um programa para avaliá-lo é perguntar como ele funciona. É necessário antes de tudo caracterizá-lo, o que não é uma tarefa fácil, pois os programas sociais incorporam grande variação em termos de componentes, estilos, pessoas e procedimentos. Torna-se, portanto, necessário adotar procedimentos, tais como: análise documental, entrevistas com decisores e implementadores das ações/políticas de órgãos governamentais, entrevistas com especialistas da comunidade científica e com os usuários.

## Conceitos básicos: um olhar contextual

No *Dicionário Aurélio*, “avaliar” significa “determinar a valia ou o valor de” ou ainda, “determinar a valia ou o valor, o preço, o merecimento, etc.; calcular, estimar”. Assim, cada pessoa realiza numerosas avaliações ao longo do dia, pois cada vez que se julga um objeto em relação a uma dimensão de valor (preço, merecimento, etc.) está sendo realizada uma avaliação.

Existem vários tipos de avaliação. [Scriven \(2001\)](#) cita alguns deles: avaliação de pessoal; estudos de política; avaliação de desempenho (no caso de estudantes, atletas, combatentes de guerra); análise/*reviewing* de restaurantes; avaliação de programa; pista de teste para veículos; avaliação do impacto ambiental; crítica a títulos de investimentos; crítica a jogos de carta e outros de azar, inclusive, o de guerra; ética normativa (isso é, aplicada); a lógica prática (argumentação) e análise de decisão aplicada.

Para esse autor, esses diversos tipos compõem o território da avaliação. No entanto, ressalta dois campos que só recentemente foram identificados e considerados de grande importância: o primeiro é o da meta-avaliação (a avaliação das avaliações) e o outro, o campo da avaliação intradisciplinar, isso é, a avaliação conduzida dentro de disciplinas.

Scriven (2001) considera que a avaliação por si só já é uma disciplina de um tipo familiar, mas não convencional. Ela faz parte de um grupo transdisciplinar de elite por pertencer ao grupo das mais notáveis, que servem a outras disciplinas, embora tenha também o *status* de condição autônoma. Nesse sentido, a avaliação pode ser utilizada não só no campo das ciências sociais, mas também no da História, da Geografia, da Ciência da Computação, da Educação, dos Negócios e Gestão, bem como das ciências mais antigas, como a Física e a Biologia. Pode-se dizer, assim, que a avaliação é uma tarefa eminentemente multidisciplinar ([FARIA, 2005](#)).

Por sua vez, [Silva \(2001\)](#) considera que a avaliação de políticas e programas sociais é um campo de estudo da pesquisa social em desenvolvimento, verificando-se sua expansão a partir dos anos 1960 nos Estados Unidos, no âmbito dos programas de combate à pobreza. Todavia, naquele contexto, [Faria \(2005\)](#) afirma que a pesquisa avaliativa era concebida predominantemente como uma ferramenta de planejamento destinada a formuladores de políticas e a gerentes do mais alto escalão. No Brasil, a expansão da avaliação de políticas públicas é registrada a partir dos anos 1980, quando os movimentos sociais passaram a demandar políticas sociais universalizadas, como um direito de cidadania. Nessa mesma época, aprofundou-se a crítica ao padrão de políticas sociais desenvolvido na América Latina e, especificamente, no Brasil.

Essas críticas referem-se, contundentemente, ao mau uso do dinheiro público e à desfocalização dos programas sociais em relação à população mais necessitada. A disseminação ainda se deve às exigências dos organismos internacionais, como o Banco Mundial e o Banco Interamericano de Desenvolvimento, os quais demandaram reforma dos programas sociais, pautada na descentralização e privatização. Nesse sentido, as instituições financiadoras incluíram a avaliação das políticas sociais como condição para seu financiamento, passando a exigir racionalidade do gasto público e do rendimento dos recursos aplicados. Desenvolveu-se então a avaliação com a finalidade de aferir a eficiência no uso dos recursos, bem como dimensionar o grau de eficácia com que os objetivos dos programas estavam sendo alcançados (FARIA, 2005).

Do ponto de vista desse mesmo autor, a avaliação de políticas públicas pode ser considerada uma atividade destinada a: (1) melhorar os resultados de uma ação cujo ciclo de vida se encerra; (2) fornecer elementos para o desenho de novas intervenções ou para o aprimoramento de políticas e programas em curso; (3) fazer parte da prestação de contas e da responsabilização dos agentes estatais. E, nesse contexto, destacam-se os *stakeholders*, que, segundo [Brasil](#)

(2002) e [Silva \(1999\)](#), são aquelas pessoas, grupos ou instituições envolvidos pela política e nela interessados, cuja influência é decisiva quanto ao sucesso do programa/projeto (atores sociais). Esses *stakeholders* estão divididos em grupos primários e secundários. Os grupos primários são aqueles mais afetados seja positivamente (os beneficiários), seja negativamente (os realocados involuntariamente); os secundários são os intermediários envolvidos no processo de prestação de serviços.

Nessa perspectiva, usam-se indicadores de desempenho para aferir os resultados obtidos, definindo esse termo como um número, percentagem ou razão que mede determinado aspecto do desempenho, com o objetivo de compará-lo às metas preestabelecidas. Diferencia, ainda, medidas de desempenho quando seus aspectos podem ser mensurados diretamente e quantificados facilmente, e indicadores de desempenho, quando não é possível efetuar tais mensurações de forma direta. Embora sejam indiretas, considera-se que essas medidas podem fornecer uma boa visão do resultado a ser medido, esclarecendo que são meras aproximações da realidade ([BRASIL, 2000](#)).

Similarmente, [Mourão \(2004\)](#) define os indicadores como medidas explícitas e objetivamente verificáveis de mudanças induzidas ou de resultados. Seriam, pois, formas de “capturar” os resultados e também de conseguir transmiti-los. Enquanto os objetivos devem responder à pergunta “O que mudou?”, os indicadores devem responder à pergunta “Como se sabe que mudou?”. Além disso, a autora denomina indicadores duros, ou medidas duras, ou, ainda, números duros, os indicadores numéricos (ex.: percentual de inadimplência apresentado pelo agente financeiro/banco) em relação aos indicadores perceptuais (ex.: percentual de satisfação indicado pelo beneficiário).

A esse respeito, a referida autora faz uma observação, relatando que indicadores perceptuais, coletados por meio da comunicação, são mais versáteis e econômicos e têm como vantagem a possibilidade de serem usados na maior parte das informações e dados necessários e como desvantagem, serem baseadas em crenças. Nesse caso, dependem da

visão que as pessoas têm do fato, da sinceridade, da memória e da boa vontade dos respondentes. Importante, também, é destacar os vieses que podem ser gerados em virtude da “desejabilidade social”, mediante a qual os pesquisados podem, conscientemente ou não, responder aquilo que acreditam que o pesquisador gostaria de ouvir, ou que é mais aceito naquela determinada cultura.

Os indicadores quase sempre são compostos de variáveis, como custo, tempo, quantidade e qualidade. Nesse sentido, o desempenho na obtenção de determinado resultado pode ser medido considerando dimensões de análise, tais como economicidade, eficiência, eficácia e efetividade. Para cada uma dessas dimensões, podem existir um ou mais indicadores ([BRASIL, 2000](#)).

[Mourão \(2004\)](#) acrescenta a essas dimensões critérios referenciais, tais como: a equidade, o impacto social, o custo-benefício, o custo-efetividade, a satisfação dos beneficiários, o acatamento das normas e das regras legais em sua implementação, entre outros. Essa autora considera que cada um desses critérios tem perguntas centrais a serem respondidas e requerem um enfoque diferente para a execução da investigação.

Além dos indicadores de uma pesquisa avaliativa, há outros conceitos a serem considerados. Na literatura, costuma-se distinguir a avaliação de políticas públicas/programas em termos de sua eficiência, eficácia e efetividade. Essa diferenciação é basicamente um recurso analítico destinado a separar aspectos marcantes dos objetivos e, por conseguinte, da abordagem, dos métodos e técnicas de avaliação ([ARRETCHE, 2001](#)).

Esses três conceitos são indicadores estratégicos para a construção de um modelo de avaliação de programas, uma vez que possibilitam a identificação e a análise detalhada das diferentes etapas que envolvem os trabalhos de operacionalização de uma política pública. Assim, a compreensão desses conceitos é de suma importância.

Por avaliação de eficiência, entende-se um esforço empregado na implementação de determinada política e os resultados alcançados. A autora considera que a avaliação de eficiência é hoje, possivelmente, a mais necessária, a mais urgente de ser desenvolvida. Grande tem sido o esforço na sofisticação dos métodos de avaliação de eficiência nos últimos anos. Isso tem ocorrido em razão de vários fatores, tais como o desperdício de recursos, a corrupção ou a incapacidade gerencial no âmbito das organizações estatais, os quais se tornam grandes entraves à utilização de recursos públicos geridos para finalidades efetivamente sociais ([ARRETCHÉ, 2001](#)).

[Belloni et al. \(2001\)](#) assinalam que eficiência diz respeito à relação entre o previsto e o realizado, com a finalidade de combinar os insumos e implementos necessários à consecução dos resultados visados. Referem-se à otimização dos recursos utilizados, funcionando como causa esclarecedora do resultado que se busca avaliar, além de ser um indicador de produtividade das ações desenvolvidas. Abrange método, metodologias, procedimentos, mecanismos e instrumentos utilizados para planejar, projetar, tratar objetos ou problemas, tendo em vista as diretrizes e determinados objetivos finalísticos.

Por avaliação de eficácia, [Arretche \(2001\)](#) entende a relação entre os objetivos e os instrumentos explícitos de um dado programa e seus resultados efetivos. A autora observa que essa avaliação pode ser feita, por exemplo, entre as metas propostas e as metas alcançadas pelo programa, ou entre os instrumentos previstos para sua implementação e aqueles efetivamente empregados.

Para [Belloni et al. \(2001\)](#), eficácia corresponde ao resultado de um processo, incluindo também a orientação metodológica adotada e a atuação estabelecida na consecução de objetivos e metas, em um tempo determinado, tendo em vista o plano, o programa ou o projeto originalmente proposto. Quando se trata de política pública, considera-se que eficácia não pode estar restrita tão-somente à aferição de resultados parciais. Ela se expressa, também, pelo grau de qualidade do resultado atingido. Em síntese, para [Varaschin et al. \(2002\)](#), um projeto é considerado eficaz quando consegue realizar seus objetivos.

Esses conceitos correspondem aos de Ribeiro et al. (1998), que consideram eficiência a execução correta das atividades programadas e eficácia a adequação das atividades e dos produtos para garantir o alcance do objetivo imediato do projeto, o que implica a verificação do grau de satisfação do público-alvo.

Por avaliação de efetividade, [Arretche \(2001\)](#) entende a relação entre a implementação de determinado programa e seus impactos e (ou) resultados, isso é, seu sucesso ou fracasso em termos de uma efetiva mudança na qualidade de vida das populações atingidas pelo programa sob avaliação. Nesse tipo de avaliação, a distinção entre avaliação e análise se torna mais clara e necessária em razão da necessidade de demonstrar que os resultados encontrados na realidade social estão causalmente relacionados àquela política particular e da necessidade de distinguir entre os produtos de uma política e seus resultados.

Para [Belloni et al. \(2001\)](#), a efetividade social refere-se às conseqüências da qualificação do trabalhador, em sua vida profissional e pessoal. Destaca-se, assim, a efetividade social como um critério de avaliação que procura dar conta dos resultados da política pública, tanto econômicos quanto sociais. Consideram, ainda, que a compreensão da efetividade social envolve indicadores de caráter macrossocial, nem sempre disponíveis quando se desenvolve avaliação de política pública, em especial, quando os resultados de uma política podem ser manifestados a longo prazo ou mediante indicadores difíceis de serem diretamente aferidos. A alternativa é combinar indicadores diretos e indiretos, uma vez que ganhos de renda ou de *status*, por exemplo, para um programa de educação, nem sempre expressam todo o impacto e as conseqüências do programa.

[Ribeiro et al. \(1998\)](#) consideram que a efetividade, em última instância, mede o impacto do projeto em relação ao seu objetivo superior. Isso significa dizer que um projeto será tanto mais efetivo quanto mais próximas estiverem suas atividades e produtos (ou metas) do objetivo superior. Consideram, ainda, que o esforço via projeto e os seus conseqüentes produtos são efetivos quando conduzem ao objetivo

superior, produzindo impacto positivo. Resumidamente, para [Varaschin et al. \(2002\)](#), o resultado é aquilo que os projetos produzem, como desempenham o seu grau de eficiência.

Apesar de esses conceitos terem sido discutidos de forma isolada, há que considerá-los interligados, relacionados. Por exemplo, um programa voltado à reforma agrária pode ter sido eficiente em suas operações, eficaz no alcance de suas metas, como é o caso de se assentar certo número de pessoas em um determinado prazo. No entanto, não ter sido efetivo, por não ter alcançado êxito quanto aos resultados, à possibilidade de redução da pobreza, à melhoria de renda e da qualidade de vida em um horizonte de tempo.

Isso vale também para a questão do crédito agrícola. Empréstimo certo volume de recursos pode significar o alcance de determinada meta de um programa, mas não quer dizer que os recursos estejam melhorando a qualidade de vida, a renda, a capacidade de produção de produtores e, conseqüentemente, estão sendo devolvidos em conformidade com o ajustado.

Além desses, o conceito de impacto no campo da avaliação de programas é um termo adotado por diversas áreas da ciência. Na psicologia organizacional, [Castro \(1999\)](#) define o impacto do treinamento no trabalho como sendo a influência exercida pelo treinamento sobre o desempenho subsequente do treinando em tarefa similar àquela aprendida por meio do programa instrucional. Nesse contexto, o impacto compreende: (1) utilização freqüente, no trabalho, dos conhecimentos adquiridos; (2) melhoria da qualidade do desempenho e (3) diminuição do número de erros cometidos no trabalho. Para a presente pesquisa, o impacto representa o resultado final, o efeito do programa na vida do usuário, em termos de melhoria da qualidade de vida, de renda, de geração de emprego, de capacidade de produção, de fixação no campo. Não basta adquirir um animal se ele não traz nenhum benefício, não gera meios de vida suficientes para ser considerado impacto.

Pode-se dizer, então, que a avaliação de um programa social tem por finalidade determinar se o programa atingiu ou não os objetivos previstos. Em outras palavras, trata-se de comparar as dimensões relevantes em dois momentos do tempo, antes e depois da intervenção, para comprovar se esta provocou a mudança esperada. Todavia, mesmo que a mudança tenha acontecido na direção prevista, isso não garante automaticamente que ela seja causada pelo programa. Isso porque as dimensões consideradas não são estáticas, podendo receber influência de uma série de fatores. Assim, o importante é determinar a causa das mudanças. Por sua vez, se não houver mudanças, cabe também identificar a causa dessa estabilidade, pois pode acontecer que o programa tenha surtido efeito positivo, mas que foi anulado pela influência de outros elementos que nada têm a ver com o programa (CANO, 2004).

O autor segue argumentando que, na avaliação, procuram-se adjetivos que indiquem o tipo concreto de avaliação que estaria sendo realizado. Daí o estabelecimento da primeira distinção clássica entre a avaliação somativa e a formativa (SCRIVEN, 1996). Avaliação somativa é aplicada para verificar se o programa social surte efeito desejado ou não, e avaliação formativa, no contexto de um programa em andamento, visa avaliar seu desenvolvimento, como pode ser melhorado. O primeiro tipo responde à pergunta: “o programa cumpre seus objetivos?”. Já o segundo responde à pergunta: “como é possível melhorar o programa?”.

Scriven (1996) defende a dicotomia formativa/somativa como uma das razoáveis maneiras de classificar as atividades de avaliação. Ele argumenta que o contexto em última instância é que determina se as atividades são uma função formativa ou somativa, se um tipo de avaliação é mais importante que o outro. No campo da avaliação social, por exemplo, entre muitas outras, como no mundo do esporte, as avaliações somativas são o foco dos grandes eventos, como as Olimpíadas e a Copa do mundo. A importância relativa é, novamente, um assunto essencialmente de contexto, não de lógica ou metodologia.

Para [Donaldson e Gooler \(2003\)](#), enquanto a condução de uma avaliação formativa é vista como desenvolvimentista, a da somativa é, às vezes, percebida como um julgamento, uma ação punitiva.

Relacionada a essa classificação, está a diferenciação entre avaliação interna e externa. A avaliação interna é desenvolvida por pessoas ou equipes que pertencem ao programa, enquanto a externa é conduzida por consultores ou instituições que não fazem parte desse programa e que são chamados especificamente com essa finalidade. As avaliações somativas são mais conhecidas como externas, e as formativas, em geral, internas ([CANO, 2004](#)).

Uma equipe externa terá maior isenção, maior distância em relação aos costumes e às normas da instituição e não se verá constrangida por fazer críticas; seu trabalho termina juntamente com a avaliação e não depende necessariamente, como seria o caso do pessoal interno, do resultado dela. Ademais, uma equipe interna poderá conhecer melhor o programa, prestar maior ajuda com sugestões durante seu desenvolvimento e, provavelmente, enfrentará menor resistência por parte dos colegas quanto ao fornecimento de informação exaustiva e veraz.

Em outras palavras, tanto as pessoas que estão sob o impacto do programa, quanto aquelas que o operacionalizam, são avaliadores em potencial. No entanto, aqueles que conseguem levantar todas as informações do programa junto ao grupo envolvido com o programa, como é caso dos consultores, tornam-se os avaliadores de maior potencial.

Ainda dentro desse prisma, [Faria \(2005\)](#) enfatiza a questão do uso da avaliação considerando quatro tipos:

(1) Instrumental, relativo ao apoio às tomadas de decisões e à busca de resolução de problemas. Nessa perspectiva, [Marinho e Façanha \(2001\)](#) enfatizam que problemas de gerenciamento de políticas e de programas sociais podem receber contribuições práticas e operacionais

das atividades de avaliação, uma vez que ela adquire condição de instrumento central e indispensável para a gestão.

(2) Conceitual, voltado para questões educativas, em que as descobertas podem alterar a maneira como os técnicos compreendem a natureza, o modo de operação e o impacto do programa que implementam. Nesse contexto, [Silva \(1999\)](#) propõe que a idéia da avaliação como instrumento de correção de rota deve ser substituída pela noção de aprendizado na implementação de uma política pública (*policy learning*). A implementação e a avaliação de políticas devem ser entendidas como testes de modelos causais sujeitos à corroboração ou ao abandono.

(3) Instrumento de persuasão, focado na obtenção de apoio para a posição que os tomadores de decisão já têm sobre as mudanças necessárias na política ou no programa.

(4) Instrumento de esclarecimento, voltado para o impacto sobre as redes de profissionais, sobre os formadores de opinião e sobre as posturas partidárias (*advocacy coalitions*), focado também nas alterações das crenças e na forma de ação das instituições, constituindo, assim, a agenda governamental. Dessa forma, o autor considera que a distinção das formas de uso constitui avanço significativo em relação à contraposição entre avaliação “somativa”, que seria orientada para julgamento (*judgement-oriented*), e avaliação “formativa”, que seria orientada para o desenvolvimento (*improvement-oriented*).

[Cozby \(2003\)](#), bem como [Posavac e Carey \(2003\)](#), apresentam diferentes tipos de avaliação. O primeiro é a avaliação da necessidade, em que se procura identificar e medir o nível de necessidades desconhecidas dentro de uma organização ou comunidade. Nesse caso, os avaliadores podem examinar o perfil socioeconômico da comunidade; o nível de problemas sociais dentro das comunidades, das agências e instituições que, no dia-a-dia, servem à comunidade. Os dados para

a avaliação de necessidades podem ser originados de levantamentos, entrevistas e dados estatísticos provenientes de órgãos públicos e outras agências.

Em uma pesquisa, [Lee et al. \(2007\)](#) estudaram as seguintes questões relacionadas à avaliação de necessidades: (1) Qual a opinião de grupos de atores sociais (estudantes e administradores de faculdade) a respeito da manutenção de serviços em termos de importância, satisfação e frequência de uso?; (2) Os grupos têm perspectivas similares ou diferentes em relação aos serviços, especialmente, em termos de discrepância entre a importância e a satisfação?; (3) Se há diferença, quais as razões que eles dão para explicar a visão deles e de outros grupos?

Uma metodologia mista com elementos da abordagem qualitativa e quantitativa foi utilizada. Para o primeiro estudo, com base na percepção dos alunos e foco na manutenção dos serviços, foi aplicado, via Web, um instrumento de 18 itens. Para os membros e administradores da faculdade, um instrumento similar foi elaborado e aplicado. Uma escala tipo Likert de 5 pontos foi usada para medir os três aspectos do serviço: importância, satisfação e frequência de uso. Verificou-se, por exemplo, na percepção dos respondentes, até que ponto o serviço foi importante (1 = discordo totalmente, 5 = concordo totalmente). A discrepância entre os escores da importância e da satisfação serviu como índice de necessidades.

Para o segundo estudo, uma pesquisa de manutenção de contato (*follow-up survey*) foi conduzida contendo questões semi-abertas para explorar a natureza das diferenças dos grupos. Técnicas de estatística descritiva e inferencial (ANOVA) foram aplicadas aos dados. Para ambos os grupos estudados, com foco no serviço, ficaram evidenciadas taxas consistentes de importância (com tendência descendente), satisfação e frequência de uso. Geralmente, os escores de importância foram mais altos, seguidos pela satisfação e frequência de uso. Nenhum serviço obteve índices altos de necessidade.

Outro exemplo desse tipo de pesquisa é apresentado por Unruh (2005), ao relatar que a participação de atores sociais, especificamente os beneficiários primários, melhora os resultados da avaliação pelo aumento da validade geral das medidas avaliativas. Essa investigação, relativa à avaliação de necessidades, foi realizada com base em cinco fóruns comunitários regionais constituídos por jovens de instalações correccionais e por provedores de serviços básicos da comunidade (N = 127), como atores sociais secundários.

Um processo de grupo nominal (*nominal group process*) foi conduzido para definir os suportes e as barreiras para o desenvolvimento do modelo de entrega dos serviços. Além disso, foram aplicadas, nesses jovens avaliadores (N = 33), entrevistas qualitativas visando descrever a percepção deles sobre as barreiras e suportes. Ambos os grupos identificaram e classificaram 12 suportes e barreiras. Temas comuns emergiram com grandes diferenças entre as informações que ajudaram a modelar o programa de entrega de serviços, de forma a permitir acomodar as necessidades de ambos os grupos: o dos jovens avaliadores e o dos atores sociais secundários.

O segundo tipo, segundo [Cozby \(2003\)](#), é a avaliação da teoria do programa em que se pretende planejar um modelo. O especialista deve basear-se em suposições válidas sobre as causas dos problemas e os aspectos racionais do programa proposto. A avaliação da teoria do programa pode envolver a colaboração de pesquisadores, de prestadores de serviço e de futuros clientes, para verificar se o programa proposto de fato atende às necessidades da população-alvo.

O terceiro tipo de questão diz respeito à avaliação dos processos ou monitoramento do programa. Enquanto está sendo implantado, os pesquisadores responsáveis pela avaliação, monitoram-no para determinar se ele está atingindo a população-alvo e se a equipe de profissionais está oferecendo os serviços planejados. A equipe pode não ter recebido treinamento adequado ou os serviços podem estar sendo oferecidos em um local indesejável ou de difícil acesso. Em resumo, o pesquisador deseja garantir que o programa esteja atingindo

seus objetivos. Isso é um fato relevante, porque não é desejável concluir que um programa seja ineficiente, quando de fato o problema reside em sua implementação. Para isso, segundo Silva (1999), devem-se empregar mais métodos qualitativos que quantitativos, ou, até mesmo, a combinação dos dois. Uma pesquisa com essa finalidade pode envolver questionários e entrevistas semi-abertas, estudos de observação e análise de registros mantidos pela equipe do programa.

Esse tipo de avaliação está inserido no contexto da presente proposta, cujo objetivo é avaliar a adimplência e a inadimplência dos agricultores familiares dos grupos A e B do Pronaf. Para isso, a avaliação de processo, nesse caso, tem a ver com a maneira como os técnicos do programa estão implementando o serviço creditício, incluindo a eficiência na comunicação verbal e não-verbal, a assimetria de informação, bem como outras ações de interação interpessoal ou de massa.

Segundo [Azevedo e Shikida \(2004\)](#), assimetria de informações refere-se à diferença entre as informações do ofertante e do demandante pelo crédito bancário. No tocante à oferta, o agente financeiro não pode conceder crédito a todo tomador interessado, uma vez que a falta de informações poderá colocar em risco seu portfólio de empréstimos. Por sua vez, os tomadores devem estar bem informados a respeito dos termos dos contratos ou diferentes documentos que estão assinando, caso contrário podem comprometer sua renda/capital por conta de uma desinformação.

A quarta questão diz respeito à avaliação dos resultados ou avaliação de impacto: “os resultados pretendidos pelo programa estão sendo atingidos?”. Para determinar se estão, o avaliador deve encontrar uma forma de medi-los para então estudar o impacto do programa sobre essa medida. Precisa-se saber também como estão os participantes do programa e como estariam se não tivessem passado por ele. [Cozby \(2003\)](#) ressalta que, idealmente, um delineamento experimental com atribuição aleatória dos participantes às condições experimentais seria planejado para responder a questões sobre resultados. No entanto, outras abordagens de pesquisa, como o delineamento

quase-experimental e o delineamento com único sujeito, são formas muito úteis de se avaliar o impacto de um programa de intervenção.

A questão final de avaliação de um programa diz respeito à avaliação da eficiência. Tendo demonstrado que um programa tem o efeito pretendido, os pesquisadores devem determinar se ele “se justifica”. Custos e benefícios devem ser ponderados. Além disso, os pesquisadores devem determinar se os recursos usados para implantar o programa poderiam ser mais bem empregados de outra forma.

### **Aspectos teóricos e metodológicos: uma breve discussão**

Apesar dos imbricados conceitos envolvidos neste campo da pesquisa, conforme visto anteriormente, [Scriven \(1998\)](#) chama a atenção para a natureza das teorias que envolvem a prática da avaliação.

Para esse autor, as teorias são conjuntos de proposições de onde se obtêm as explicações e a integração de componentes. Ainda que obtidas de relatos, não são teorias de operações, uma vez que não proporcionam nenhuma explicação do ajuste ou da unificação lógica das operações. São simplesmente conjuntos de instruções para juntar partes de um processo. Nem são conjuntos simples de observações sobre as conexões entre componentes das teorias de avaliação. No máximo, são conjuntos parciais de proposições de nível intermediário que têm de ser derivados de quaisquer coisas que dêem a entender que são teorias. No entanto, elas podem ser suficientes para identificar a(s) causa(s) da(s) falha(s) do sistema avaliado.

Scriven (1998) acrescenta, ainda, que as teorias proporcionam explicações gerais e econômicas, com fins de simplificação conceitual e (ou) a explicação de um fenômeno, e envolvem mais um grau de especulação que observações diretas. Em sua opinião, a predição é só um ganho a mais, não uma condição necessária para que alguma coisa seja uma teoria, até mesmo no caso de uma boa teoria. Existem muitas disciplinas respeitáveis, de Geologia a História, em que as melhores teorias não geram predição.

O mesmo autor reforça a idéia da tão conhecida frase atribuída a Kurt Lewin: “Não existe nada tão útil como uma boa teoria”. Para ele, a primeira coisa que uma teoria deveria facilitar é a ordenação dos componentes de um modelo por meio de alguns critérios de demarcação. Ela deveria permitir a diferenciação da avaliação de outros tipos de investigação, tais como a predição, a análise de dados, a diagnose, a pesquisa, a explicação, a descrição, a generalização e a classificação.

A formatação da teoria minimalista (*minimalist theory*) se inicia com questões de demarcação. Por conseguinte, inclui em seu modelo um conjunto de nove elementos básicos, organizados em passos ou componentes (às vezes implícitos) dentro da lógica da avaliação, como: (1) a determinação de até que ponto qualquer das três propriedades – mérito/qualidade, valor/custo da efetividade ou significância/importância – pode ser atribuída à entidade que está sendo avaliada; (2) todas as conclusões devem ser expressas em termos de quatro atributos: a posição de algo em termos de qualidade ou importância (*ranking*), o grau ou nível de avaliação (*grading*), a pontuação (*scoring*) e a divisão (*apportioning*); e (3) as recomendações e as explicações nem sempre são deduzidas das conclusões ([SCRIVEN, 1998](#)).

Por fim, o autor apresenta as seguintes características dessa teoria: (1) integra diferentes campos da ciência; (2) provê um mapa de território, baseando-se na definição do assunto, com demarcação da avaliação de muitos outros campos adjacentes que estão confusos; (3) provê uma listagem concisa dos conceitos abstratos centrais do campo de estudo; e (4) implica vários resultados desses conceitos de real valor para os campos aplicados.

[Donaldson e Gooler \(2003\)](#) apresentam a teoria da avaliação dirigida, explicando que o foco dessa teoria é a compreensão total da natureza do programa, dos objetivos verdadeiros e do contexto da avaliação, em um esforço para se projetar com maior precisão a avaliação dentro das limitações práticas. Para isso, defendem a utilização de métodos quantitativos, qualitativos e mistos, dependendo da situação de

avaliação. Nesse sentido, consideram que a teoria de programa é um conjunto de proposições em relação ao que vai dentro de uma caixa preta durante a transformação do que entra (*input*) para o que sai (*output*), isso é, como uma situação é transformada para melhor, via tratamentos (*inputs*). É altamente desejável que a teoria do programa esteja enraizada ou pelo menos seja consistente com a teoria das ciências sociais ou comportamentais.

No contexto da teoria da avaliação dirigida participativa, [Donaldson e Gooler \(2003\)](#) relatam que essa abordagem fundamenta-se no desenvolvimento de teorias de programas para cada tipo e usam testes empíricos como retroalimentação para guiar o desenvolvimento dos programas. Cada teoria é baseada na experiência dos atores sociais de como esses programas parecem funcionar. Pesquisas de avaliação anteriores e de trabalhos teóricos e empíricos servem de guia de modelagem para projetos de avaliação desenvolvidos especificamente para responder, tão rigorosamente quanto possível, às perguntas-chave de pesquisa, dadas as limitações do contexto de avaliação. Uma vez que fatores ambientais apresentam grande potencial para confundir as estimativas dos efeitos do programa, essa abordagem cartográfica conceitual ajuda a identificar e examinar as fontes de variância e isolar os efeitos desses fatores sobre o impacto do programa.

Nessa perspectiva, [Silva \(1999\)](#) ressalta dois tipos de teoria utilizados nos estudos e pesquisas de avaliação: a Teoria do Programa, cujos princípios se baseiam na idéia de que todo programa traz em si uma teoria e que a avaliação representa o teste; e a Teoria de Implementação, fundamentada na premissa de que o programa funciona como planejado e que os resultados esperados são atingidos. Com base nessa idéia, deve-se entender em que medida o programa funciona ou não como esperado, possibilitando verificar como os obstáculos no plano da implementação distorcem as intenções originais dos formuladores.

Em avaliação de programa, é relevante saber não apenas o que se propõe atingir, mas como se pretende alcançar os objetivos. Todo

programa baseia-se em premissas ou hipóteses testáveis sobre os meios necessários para se chegar a determinados fins. Por exemplo, um programa de melhoria de desempenho de servidores públicos é construído com base no argumento de que o aumento dos salários conduz à melhoria do desempenho no trabalho. O interesse em desvendar a relação, ou possíveis mecanismos causais, entre salários e desempenho, passa a constituir o que se chama de teoria do programa.

Para conhecer ou formular a teoria implícita de um programa, [Silva \(1999\)](#) recomenda que se deva compreender as idéias e as premissas gerais que ligam os *inputs* do programa aos objetivos pretendidos. A teoria de um programa consiste em fornecer uma explicação das seqüências causais que articulam meios e fins gerados pelo programa ou mesmo um modelo plausível de como é esperado que o programa funcione.

No que diz respeito à teoria da implementação, o referido autor argumenta que esse tipo de teoria não lida com a análise dos processos intermediários entre os serviços e os resultados alcançados, mas com a questão da oferta de serviços gerados pelo programa. A teoria da implementação permite focar as atividades específicas e verificar em que medida cada uma das atividades funciona como planejado.

A construção de uma teoria do programa inclui as seguintes atividades: (1) os *inputs* do programa; (2) as atividades operacionais, as quais representam o modo específico pelo qual o programa é implementado; (3) resultados obtidos, que representam a cadeia de respostas para cada atividade do programa e (4) os resultados esperados com a implementação (SILVA, 1999).

Os aspectos teóricos requerem um estudo acurado dos aspectos metodológicos e, nesse sentido, a avaliação de programas sociais teve como um de seus principais precursores Donald Campbell, que apresentou uma metodologia para avaliar intervenções sociais considerando desenhos experimentais, quase-experimentais e correlacionais ligadas à pesquisa na área de ensino ([CAMPBELL;](#)

[STANLEY, 1979](#)). Segundo [Cano \(2004\)](#), a posição de Campbell, de origem experimentalista, foi criticada por avaliadores que defendiam uma abordagem mais qualitativa e mais centrada na visão dos atores. As técnicas dessa nova corrente tinham cunho mais antropológico, como a etnografia ou a observação participante. Além disso, o estudo de caso usando uma abordagem qualitativa era outra técnica freqüentemente utilizada.

A lógica dos trabalhos de Donald Campbell não está relacionada exclusivamente às pesquisas quantitativas, ainda que tenha sido posta em prática pela maioria dos pesquisadores “quantitativistas”. Os mesmos delineamentos de pesquisa poderiam ser implementados usando técnicas qualitativas. Quando um avaliador escolhe as instituições ou locais concretos onde realizará estudos de caso a partir de entrevista em profundidade, ele considera a possibilidade de fazer inferências causais (validade interna) e de generalizar o universo de interesse (validade externa). Do ponto de vista reflexivo, a lista de ameaças aos diversos tipos de validade pode também ajudar o pesquisador que utiliza uma abordagem qualitativa a estabelecer uma estratégia que melhore o poder persuasivo de seus resultados.

Cano (2004) alerta para o fato de que, em se tratando de uma avaliação de intervenção social, é relevante selecionar o delineamento de pesquisa para garantir a comparabilidade dos grupos envolvidos, o controle de variáveis intervenientes e a superação dos problemas de seleção diferencial. [Dugard e Todman \(1995\)](#) assinalam que o efeito de um programa é considerado pelas medidas anteriores e posteriores, demonstrando a grandeza daquilo que o programa sugere ter realizado. A relação causa-efeito (efetividade) igualmente deve ser alvo de investigação. Tal autor afirma que a relação causa-efeito remete ao fato de o resultado obtido ter sido ocasionado pela ação do programa avaliado.

Freqüentemente, nos estudos de avaliação, têm sido utilizados os métodos de experimentos randomizados, incluindo a mensuração de variáveis relevantes de pelo menos dois grupos equivalentes: um que

tenha recebido influência direta do programa (grupo experimental) e outro não exposto (grupo controle), visando entender como e de que modo o programa implementado traz resultados para os seus beneficiários, ou ainda técnicas como o do grupo focal. Outros métodos podem ser utilizados, tais como os estudos de caso, estudos *ex-post*, séries temporais, estudos de correlação temporal, estudos de correlação, entre outros. ([SILVA, 1999](#)).

[Mourão \(2004\)](#) enfatiza que a maior parte dos programas produz resultados em médio e longo prazos, o que demanda delineamentos de avaliação longitudinais. Nesse sentido, a autora cita dois tipos: (1) quando os grupos-alvo devem ser comparados tão rigorosamente quanto possível aos grupos de controle; e (2) quando não havendo grupo controle, a informação sobre o comportamento pós-intervenção pode ser comparada com as medições de um período anterior ou com os resultados e conjecturas que teriam ocorrido sem a intervenção. Autores como [Campbell e Stanley \(1979\)](#) e [Cano \(2004\)](#) propõem delineamentos experimentais ou quase-experimentais para comparar um ou mais grupos experimentais ou de tratamento, com grupo controle. Mourão (2004) pondera, ainda, o caso de não ser possível a utilização de uma técnica estritamente experimental, em que todos os desenhos de análise de impacto deveriam ser empregados para comparar os resultados da intervenção com alguma estimativa do que tem ocorrido ou com o que ocorrerá na ausência da intervenção.

Campbell e Stanley (1979), autores do clássico livro *Delineamentos experimentais e quase-experimentais de pesquisa*, apresentam e discutem diferentes desenhos de pesquisas de avaliação: (1) delineamentos pré-experimentais (estudo de um único caso sem controle; pré-teste e pós-teste aplicados a um grupo e comparação de grupo estático); (2) delineamentos experimentais autênticos (pré-teste e pós-teste aplicados a grupo experimental e a grupo de controle aleatórios; o delineamento de quatro grupos de Salomon e o delineamento com grupo de controle e apenas pós-teste); (3) delineamentos quase-experimentais (o experimento de série temporal; o de amostras temporais equivalentes; o de materiais equivalentes;

o delineamento com grupo de controle não-equivalente; esquema contrabalançado; entre outros); (4) delineamentos correlacionais e *ex post facto*.

[Campos \(2001\)](#) diferencia, de forma simples, o delineamento experimental, ou aquele que, necessariamente, possui controle de variáveis, manipulação de variáveis e amostra randomizada, do delineamento quase-experimental, aquele que, necessariamente, possui uma ou duas das condições do delineamento experimental. Apresenta ainda dois tipos de pesquisa: o longitudinal, em que se avalia a mesma variável, em um mesmo grupo de sujeitos, ao longo de um período de tempo; e o estudo transversal, aquele no qual se estuda a mesma variável em grupos em momentos diferentes.

Nessa perspectiva, a abordagem correlacional, apesar de não indicar necessariamente causalidade, é relativamente de baixo custo operacional e pode fornecer um levantamento preliminar de hipóteses. As que sobreviverem a esse levantamento poderão ser testadas por meio de uma manipulação experimental mais cara ([CAMPBELL; STANLEY, 1979](#)).

Tendo sido selecionado o método analítico adequado ao tipo de questão, segundo [Silva \(1999\)](#), os estudos normalmente partem de um conjunto de teorias sobre o funcionamento do programa e hipóteses, as quais são empiricamente testáveis. Quando dotadas de instrumentos teóricos e metodológicos adequados, as pesquisas de avaliação possibilitam ampliar a compreensão de como e em que medida os objetivos desejados por uma política, programa ou projeto estão sendo alcançados e, sobretudo, se essas são passíveis de serem reaplicadas em qualquer programa e em qualquer escala.

[Cronbach \(1982\)](#) e Silva (1999) enfatizam que o objetivo central da avaliação de programa difere dos objetivos básicos da pesquisa social, uma vez que seus objetivos são mais apropriados aos diferentes contextos políticos e institucionais, os quais estão mais diretamente ligados à prática e a processos de tomada de decisão. A estratégia da

pesquisa de avaliação requer, por isso, uma consideração especial. A logística é a mesma em ambos os casos, mas a utilização da logística requerida depende do contexto, do objetivo e da expectativa de retorno. Muitas recomendações apropriadas para programas de longa duração referem-se ao método científico como o mais adequado para se fazer a avaliação.

Esses autores consideram, também, que as avaliações não deveriam seguir apenas uma logística, em virtude de erros que estão fora do escopo científico. Para qualquer estudo de avaliação, muitos projetos bons podem ser propostos, mas nenhum é perfeito. Nesse sentido, recomenda-se uma fase de menor controle, o estudo piloto como forma inicial de condução das pesquisas e o estudo exploratório, como estágio de alto controle de pesquisa. Em princípio, o trabalho exploratório deveria liderar a proposição científica para confirmar os dados, considerando uma ação experimental e tratamento controlado.

### **Avaliação de planos, programas e projetos: uma diversidade de métodos e modelos**

Nesse campo de pesquisa, diferentes modelos de avaliação podem ser encontrados na literatura. Entre eles, o tipo *marco lógico* como um modelo analítico voltado para a formulação, a execução, o acompanhamento e a avaliação de programas ou de projetos governamentais. É um modelo elaborado pelo Banco Interamericano de Desenvolvimento (BID), como um instrumento de planejamento e utilizado pela Agência Brasileira de Cooperação, órgão do Ministério das Relações Exteriores na análise de projetos submetidos a sua apreciação ([BRASIL, 2001](#)).

Nesse modelo, os programas, os planos ou os projetos são tratados como conjuntos de processos organizacionais, logicamente estruturados, que devem facilitar o alcance dos objetivos. A estrutura lógica deve conter, sucessivamente: a finalidade (impacto da ação governamental/benefícios almejados); o objetivo (resultados diretos associados à geração de bens e serviços); os produtos (bens e serviços ofertados/metapas fixadas) e as atividades (tarefas que devem ser executadas para a geração de bens e serviços).

Segundo [Brasil \(2001\)](#), além de ordenar os processos organizacionais, deve-se descrever cada processo/nível (finalidade, objetivo, produtos e atividades) em termos de: (1) definir indicadores práticos, mensuráveis e independentes, que devem ser expressos em termos quantitativos, qualitativos e temporais; (2) indicar as fontes que contêm dados sobre os indicadores (meio de verificação); (3) identificar fatores exógenos (pressupostos) que podem afetar o objeto da auditoria em cada nível, avaliando os riscos existentes e formulando ações alternativas.

No campo da saúde pública e privada, bem como no das organizações de serviços humanos, existem dois desafios a serem enfrentados: um em relação à medição de resultados e o outro referente à gestão por resultados. Gestores e avaliadores de programa têm tentado resolver esses desafios por meio de técnicas e estratégias, como o gerenciamento por qualidade total e a utilização de técnicas melhoradas de avaliação ([SCHALOCK; BONHAM, 2003](#)).

Apesar desses esforços, três dificuldades são tipicamente encontradas e envolvem a falta de: (1) um modelo de avaliação de programa e métodos de medição que claramente delineiem os resultados da organização e dos indivíduos; (2) um modelo lógico de programa que ajude os gestores de programa a ver as relações entre *input*, processos e resultados (de curto e longo prazo) e os papéis-chave protagonizados pela retroalimentação (*feedback*) formativa e pelas variáveis contextuais na gestão dos resultados; e (3) um mecanismo para administrar os resultados incluindo a retroalimentação para os provedores de serviços, a melhoria da qualidade dos serviços e do desempenho-padrão.

Nessa mesma linha de trabalho, [Renger e Hurley \(2006\)](#) apresentam os processos de criação de um modelo lógico usando a abordagem dos modelos lógicos de ensino (*Approach to Teaching Logic Models - ATM*). Simplesmente, consideram o modelo lógico, como em uma tabela, uma lista de *inputs*, atividades, produtos (*outputs*), resultados e impactos e ressaltam a abordagem ATM. Para eles, essa abordagem permite aos usuários declarar explicitamente por que certas atividades estão sendo desenvolvidas, por que certos resultados são esperados e como esses resultados estão sendo alcançados.

Os referidos autores explicam que o desenvolvimento de modelos lógicos, via abordagem ATM, se faz por meio de três passos:

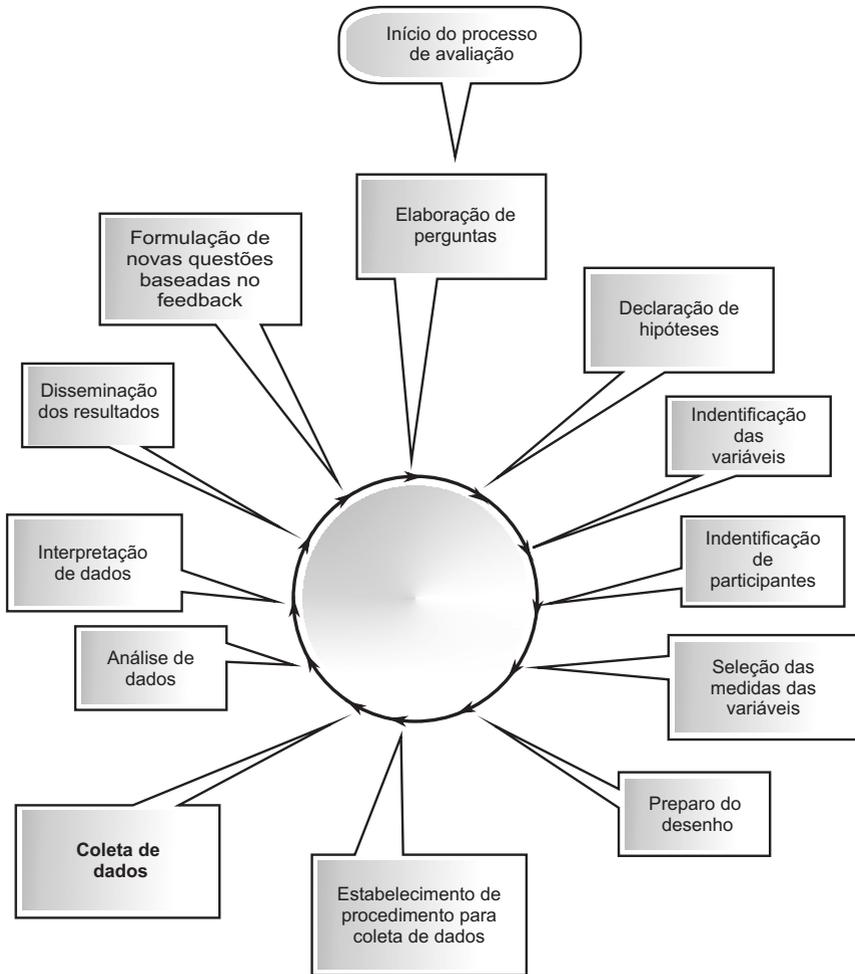
(1) Identificação das condições antecedentes: definição do problema; planejamento e condução de entrevistas; integração das entrevistas individuais em um mapa resumido; verificação das relações lineares entre os componentes do modelo (condições antecedentes e conseqüentes); complementação desse passo com fontes de informação, como a aplicação de pequeno número de entrevistas (10-12) e (ou) a revisão de literatura para complementar o mapa das condições antecedentes e a apresentação da importância desse passo na modelagem lógica.

(2) Priorização das condições antecedentes e desenvolvimento de estratégias: priorização dos processos por meio de uma simples reunião com os *stakeholders* com duração de, no máximo, duas horas; desenvolvimento de estratégias-alvo para uma ou duas condições antecedentes usando abordagem múltipla e relato da importância desse passo na modelagem lógica.

(3) Avaliação das estratégias para medir os impactos e os resultados dos programas: declaração dos objetivos do programa conectados visualmente às condições antecedentes; delimitação do plano de avaliação e relato da importância do passo três na modelagem lógica. Os autores esclarecem que o impacto do programa refere-se a um resultado imediato, enquanto o resultado em si (*outcome*) diz respeito ao resultado final, mais a longo prazo.

[Renger e Hurley \(2006\)](#) consideram que a abordagem ATM, no âmbito da modelagem lógica, é projetada para apoiar as agências no planejamento e na avaliação de seus programas.

Para [Cone \(2001\)](#), a atividade de avaliação começa com a elaboração de perguntas/questões ([Fig. 2](#)).



**Fig. 2.** Ciclo da avaliação de programa.

Fonte: [Cone \(2001\)](#).

As questões são como os problemas de pesquisa, pois iniciam um ciclo de atividades e terminam com questões extras. Existem maneiras boas e ruins de se expressar as questões de pesquisa. A boa é aquela em que a questão é respondível. O mesmo ocorre com a prática da avaliação. Para isso, deve-se especificar mais a questão, ou seja, deve-se deixá-la tão abrangente quanto possível.

Quando as questões de avaliação são bem elaboradas, fica mais fácil transformá-las em hipóteses. As hipóteses relatadas para os serviços quase sempre envolvem relações causais entre as variáveis, e a direção do relacionamento é normalmente mais clara. Frequentemente, o serviço (variável antecedente) é visto, em algum aspecto, como causa de mudança do funcionamento do programa (variável conseqüente).

O processo de definição da variável conseqüente envolve a colaboração entre o beneficiário e o avaliador, enquanto a definição das variáveis antecedentes pode vir da literatura, embora sua implementação envolva uma intervenção colaborativa sobre a situação do beneficiário. É importante que as variáveis antecedentes estejam claramente definidas para saber, com confiança, o que é responsável pelas mudanças benéficas ou pelas falhas na produção.

Adotando perspectiva semelhante, [Silva \(1999\)](#) propõe três tipos de variáveis para a análise de um modelo: as dependentes, as independentes de ordem estrutural e as de ordem político-estrutural.

Como variáveis dependentes, o autor sugere as medidas de desempenho dos programas, seja no plano de execução orçamentário-financeira, seja no plano da realização das metas e dos processos de implementação. As variáveis dependentes são as medidas de avaliação da eficácia dos programas: eficácia objetiva (metas propostas e atingidas); eficácia processual (meios e a metodologia de implantação) e o grau de satisfação dos beneficiários, dos agentes implementadores e dos gestores, tendo como indicador as opiniões dos entrevistados com base em questões objetivas.

No que se refere às variáveis independentes de ordem estrutural, Silva (1999) enfatiza aquelas que explicam as variações no desempenho dos programas, por exemplo, as que dizem respeito às diferenças e heterogeneidades dos municípios e da rede de implementação de programas sociais. Como variáveis independentes de ordem político-institucional, o autor assinala aquelas relacionadas aos atributos das organizações e dos atores ou agentes diretamente envolvidos no processo de implementação dos programas.

A clarificação das variáveis envolvidas na avaliação é essencial para selecionar as medidas de intervenção apropriadas. Para isso, [Cone \(2001\)](#) ressalta a análise dos objetivos final e instrumental. O objetivo final, também denominado por Scriven (1996) de **somativo**, refere-se ao que o beneficiário quer de diferente após o recebimento dos serviços. O objetivo instrumental, também denominado por [Scriven \(1996\)](#) de **formativo**, refere-se ao pensamento do avaliador no que deve acontecer para que a mudança ocorra.

Para avaliar a eficácia do serviço, o avaliador terá de medir esses dois tipos de objetivos. As medidas dos objetivos instrumentais tendem a ser mais específicas e focar a mudança das variáveis hipotetizadas, afetando os objetivos finais. Pode-se dizer que, para avaliar a eficácia de programas, como o Pronaf, é necessário verificar se seus objetivos estão sendo alcançados, o que falta e como fazer para melhorá-la.

Uma vez que já se conhecem as variáveis envolvidas na intervenção e suas medidas, pode-se passar para a explicação da relação entre as variáveis. Um simples delineamento comumente utilizado envolve a avaliação do cliente antes e depois do tratamento, além do exame das mudanças dos objetivos finais. O desenho da pesquisa depende da natureza de relação das hipóteses envolvidas.

Quando as medidas e o delineamento já foram selecionados, é o momento de aplicar os procedimentos para a coleta de dados. Coletar dados específicos para os objetivos instrumentais é mais desafiante que coletar dados para os objetivos finais, pois requer mais frequência e um trabalho intensivo, além de depender da confiança no participante para providenciar dados.

A intervenção pode começar deste ponto: a questão está clara, as hipóteses testáveis são disponíveis, as variáveis são conhecidas, o beneficiário está descrito, as medidas das variáveis estão disponíveis, o delineamento está à mão e os procedimentos para a coleta de dados estão no lugar. Esses trabalhos preliminares ocorrem antes de qualquer intervenção significativa.

A análise de dados, embora pareça como um passo distinto na [Fig. 2](#), ocorre concomitantemente com outras atividades. A análise de dados organiza as informações de maneira que eles possam ser interpretados. Maneiras formais de realizar isso incluem generalização e criação de gráficos demonstrativos dos dados e das estatísticas. Os gráficos podem ser úteis, pois mostram tendências e padrões sobre os quais o avaliador pode agir e fazer ajustamentos apropriados. Mesmo os gráficos de barra servem para mostrar muitas mudanças ocorridas.

Nessa perspectiva, [Silva \(1999\)](#) indica as análises: (1) univariada e o diagnóstico do programa e de sua estratégia de implementação; (2) inferencial de dados quantitativos, incluindo, por exemplo, técnicas de análise de *clusters*, análise de correspondência simples e múltipla, bem como a análise de regressão logística para testar as associações mais significativas entre as características dos agentes e das organizações implementadoras e o desempenho dos programas; (3) a análise dos dados qualitativos extraída de entrevistas e do grupo focal.

[Cone \(2001\)](#) alerta para os cuidados que devem ser tomados na interpretação dos dados em prol dos esforços da avaliação. É importante estar certo de que os dados foram entendidos completamente antes de divulgá-los.

A partir do momento em que as implicações dos dados foram completamente compreendidas, deve-se dividi-las com outras pessoas. Quando se discutem as constatações com os beneficiários, é importante colocar em perspectiva as mudanças e responder às suas questões. Interpretar os dados para os beneficiários é o primeiro passo na disseminação dos resultados e a sua reação indicará o que deve ser trabalhado com os futuros beneficiários. Com o seu consentimento, devem-se mostrar, na prática, os resultados do trabalho para os outros e verificar a reação deles. Divulgar o objetivo e a informação comparativa pode favorecer diferentes tipos de impressão em termos de futuro referencial.

As reações daqueles com quem se divide as informações poderão provocar mudanças que servem de referência para trabalhos futuros com beneficiários similares. As mudanças baseadas nas reações dos outros levam a novas questões e posição para recomeçar a avaliação de um ciclo.

Esses passos, sugeridos por [Cone \(2001\)](#), representam uma visão global e sintética do trabalho de avaliação de programa. No caso da avaliação do Pronaf, que exige, constantemente, levantamento de questões de pesquisa e medidas de variáveis, torna-se um importante referencial teórico para o estabelecimento de rotinas e rumos para o trabalho de avaliação.

Nesse mesmo contexto, a avaliação de programas, bem como de projetos, possui algo em comum: o foco ou o objeto de interesse, isso é, todas as informações convergem para o julgamento do objeto de pesquisa e, nesse sentido, diversos métodos podem ser empregados. [Ribeiro et al. \(1998\)](#) apresentam um modelo de avaliação de projetos, o qual se mostra bastante sugestivo para o desenvolvimento de outros modelos, inclusive, o de programa ([Fig. 3](#)).

O método apresentado por Ribeiro et al. (1998) baseia-se no marco lógico do sistema denominado *logical framework* ou *logframe*, uma ferramenta de identificação, elaboração, acompanhamento e avaliação de projetos inseridos no contexto mais amplo de prioridades, tanto institucionais como de programas estaduais, macrorregionais e nacionais ou que contenha algum componente de desenvolvimento. O método pode ser usado por pesquisadores ou administradores não só para identificar como também para elaborar projetos e programas, rever o andamento deles e conferir se os objetivos foram atingidos. É particularmente útil para o planejamento das atividades, dos recursos e dos insumos necessários para a consecução dos objetivos de um projeto ou de programas.

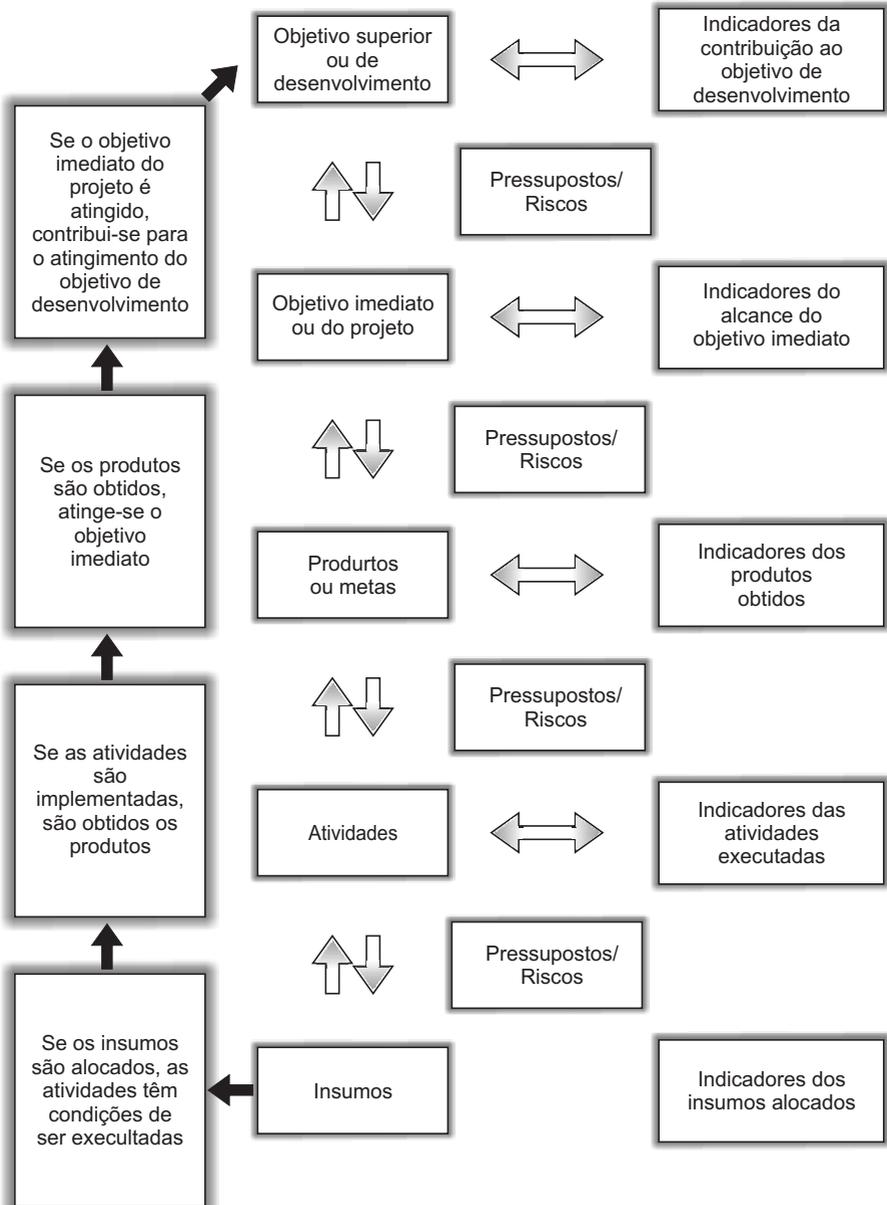


Fig. 3. Estrutura lógica dos componentes de um projeto.

Fonte: [Ribeiro et al. \(1998\)](#).

Os mesmos autores definem como objetivo superior ou de desenvolvimento algum objetivo geral da organização ou de programas de governo nas esferas municipal, estadual ou federal, para as quais o projeto poderá contribuir, devendo orientar a formulação de um ou mais objetivos imediatos. Em termos de abrangência de resultados, o objetivo de desenvolvimento está associado ao impacto do projeto fora da instituição, além do público-alvo.

O objetivo imediato, às vezes denominado de objetivo do projeto, refere-se ao objetivo que se espera alcançar perante o público-alvo, durante o período previsto para sua execução (horizonte do projeto). Deve(m) estar relacionado(s) ao objetivo superior para o qual o projeto está voltado e se propõe a contribuir para a sua consecução.

Os produtos representam as situações desejadas que se espera alcançar, no decorrer e no final da implementação do projeto. A obtenção dos produtos deverá levar ao alcance dos objetivos imediatos. Entretanto, produtos não devem ser confundidos com atividades. Em termos de abrangência, os produtos constituem resultados das atividades constantes no plano da implementação, ou seja, o esforço do projeto.

As atividades representam o esforço a ser desenvolvido por meio de ações diversas para que os produtos sejam obtidos. Podem ser expressas para cada produto especificado ou para um conjunto dos produtos, de acordo com as características do projeto.

O detalhamento das atividades possibilita a quantificação dos *insumos* necessários. A transformação dos insumos em produtos é feita por meio da execução das atividades.

O plano de implementação refere-se ao conjunto de produtos, atividades e insumos. Do ponto de vista de [Silva \(1999\)](#), a implementação diz respeito a um conjunto de atividades voltado para o alcance dos objetivos. Há que se considerar que a implementação

de políticas ocorre em um ambiente caracterizado por contínua mutação, alheia à vontade dos implementadores. Recursos, prioridades e a influência relativa dos agentes encarregados da implementação freqüentemente mudam, além dos interesses e a influência dos atores sociais, que podem, também, mudar entre o momento da formulação e o da implementação.

Os indicadores são expressões qualitativas e quantitativas de objetivos, produtos e atividades.

Os pressupostos, que não devem ser confundidos com pré-requisitos, representam a explicitação dos fatores, fora do controle do projeto, que podem influenciar o atingimento de seus objetivos, a execução das atividades e o alcance dos produtos. Se tais pressupostos não forem atendidos, o projeto poderá não alcançar os objetivos.

Os riscos representam as ocorrências anormais exógenas ao projeto, cujo acontecimento pode afetar o seu desenvolvimento e o alcance dos seus resultados, como fatores anormais de clima e mercado.

Similar a esse método, [Marinho e Façanha \(2001\)](#) apresentam um diagrama como modelo de funcionamento de um programa social. Na visão deles, de modo geral, existe um sistema de entradas e saídas, em que os recursos (entradas) são transformados em resultados (saídas). Os recursos são patrimoniais, materiais, humanos, gerenciais, informacionais, decorrentes de capacitações acumuladas e herdadas da história do programa, além dos orçamentários e financeiros. Por sua vez, os resultados não se referem exclusivamente a produtos e serviços finais ou terminados. Dizem respeito a realizações e, assim sendo, o desenho e a implementação de estratégias são entendidos como realizações. As estratégias são pontos de partida da gestão que se orienta por resultados e realizações.

A retroalimentação do sistema acentua a natureza dinâmica do funcionamento do programa, representando, assim, a perspectiva do monitoramento e do fomento, que visa verificar as realizações

alcançadas quanto aos resultados esperados, bem como ampliar os conhecimentos e recursos.

Considerando a avaliação centrada em algum aspecto de relevância do programa, [Sridharan et al. \(2007\)](#) avaliaram, por meio de planos estratégicos, a sustentabilidade de iniciativas de programas comunitários. Mediu-se, nessa pesquisa, até que ponto a sustentabilidade foi incorporada ao processo de planejamento estratégico completo. Para desenvolver uma medida de sustentabilidade, primeiro foi definido esse construto, para o qual diferentes tipos de definições foram encontrados na literatura. Entre elas, relatam uma definição em que a sustentabilidade é considerada como “o poder ou a capacidade dos programas responderem continuamente às questões identificadas da comunidade. Um programa sustentado mantém um foco consoante com suas metas e objetivos originais, inclusive com os indivíduos, as famílias e as comunidades que originalmente eram previstos”.

Para o processo de avaliação, foi utilizado um delineamento envolvendo métodos mistos. Três avaliadores analisaram 19 planos estratégicos por meio de um formulário de 36 itens relacionados a 6 construtos: componente geral; princípios de estratégia abrangente; risco e avaliação de fontes de recursos; metas concretas; recomendações e sustentabilidade.

O componente geral foi usado para medir a adequação do plano; para isso, foi utilizada uma escala contendo três itens. Os princípios de estratégia abrangente serviram para avaliar se o plano refletiu o entendimento dos princípios da estratégia completa; uma escala de três itens foi usada nessa medição. A análise do risco e a avaliação de fontes de recursos foram utilizadas para medir o rigor e a abrangência do risco, bem como o processo de avaliação dos recursos. Uma escala de sete itens foi empregada nesse caso. As metas concretas foram utilizadas para verificar a clareza e a estratégia abrangente das metas planejadas. Uma escala de cinco itens foi usada para medir as metas concretas. As recomendações/estratégias propostas serviram de base

para medir o rigor e a abrangência das recomendações e estratégias propostas e uma escala de dez itens foi usada para medir o referido construto.

A sustentabilidade foi medida por meio de oito itens, entre eles podem ser citados: (1) existe um plano e um cronograma de coleta de dados contínuos para avaliar o progresso em face das metas e da avaliação dos programas; (2) mecanismos de comunicação entre os membros (atores sociais que estiveram envolvidos durante o processo de planejamento) para a fase de implementação estão claramente especificados; (3) os processos e procedimentos são estabelecidos para assegurar que as agências cumprem suas responsabilidades.

O coeficiente de confiabilidade (alfa de Cronbach) e a precisão entre avaliadores foram calculados para cada construto usando o coeficiente de correlação *Shrout-Fleiss interclass*.

[Sridharan et al. \(2007\)](#) utilizaram uma escala de 6 pontos para os 36 itens (0 = nenhum dos componentes do plano é adequado; 1 = aproximadamente menos de 20 % desses componentes do plano são adequados; 2 = aproximadamente de 20 % a 40 % desses componentes do plano são adequados; 3 = aproximadamente de 41 % a 60 % desses componentes do plano são adequados; 4 = aproximadamente de 61 % a 80 % desses componentes do plano são adequados; 5 = aproximadamente de 81 % a 100 % desses componentes do plano são adequados). Verificaram que a sustentabilidade não foi explicada durante a fase de planejamento dessa iniciativa e que a implementação de programas e a sustentabilidade não são fases distintas e sucessivas, mas processos concomitantes.

Sem perder de vista esse foco de análise, [Mourão \(2004\)](#) avaliou o impacto de um programa público de treinamento no trabalho e na geração de emprego para pessoas com deficiência mental. Para isso, fez uso de um modelo teórico cujas variáveis critério foram estabelecidas em três níveis de impacto: (1) no comportamento no cargo; (2) na mudança organizacional e (3) no valor final.

O delineamento foi quase-experimental com três grupos: o grupo experimental principal (formado por instituições que enviaram pelo menos quatro funcionários para o treinamento); o grupo experimental secundário (com instituições que enviaram apenas o coordenador) e o grupo de controle (com aquelas que não enviaram participantes ao treinamento). A pesquisa foi realizada durante dois anos, com coleta de dados antes e depois da realização do tratamento (o programa de treinamento).

A avaliação do impacto no comportamento no cargo foi realizada com base em instrumentos já validados de auto e heteroavaliação. A avaliação do impacto na mudança organizacional e no valor final foi feita por meio dos indicadores duros: aumento no número de oficinas de formação profissional e no número de aprendizes (indicando mudança organizacional) e aumento no número de pessoas com deficiência mental empregadas (indicando valor final).

Em relação ao instrumento de medida, no que diz respeito ao que foi aprendido no treinamento para a vida pessoal e na comunidade, derivado da autopercepção do pesquisado, foram utilizados 15 itens e uma escala de 5 pontos (1 = discordo totalmente da afirmativa; 2 = discordo um pouco da afirmativa; 3 = não concordo, nem discordo da afirmativa; 4 = concordo com a afirmativa e 5 = concordo totalmente com a afirmativa). Em relação às variáveis de treinamento, estas foram mensuradas com base em três itens ancorados na escala de 1 a 5 (1 = nada adequados ao meu trabalho, ..., 5 = bastante adequados ao meu trabalho).

Os dados coletados dos questionários foram submetidos às análises estatísticas descritiva e inferencial. Para a verificação do impacto do treinamento no comportamento no cargo, considerando quatro variáveis preditoras, foram empregadas análises de regressão múltipla (seqüencial). Os impactos do treinamento nos níveis de mudança organizacional e valor final foram avaliados por meio do pré e pós-teste e amostra dividida em três grupos: experimental principal, experimental

secundário e de controle. Para a análise da ocorrência desses impactos, [Mourão \(2004\)](#) empregou quatro procedimentos: Anova dos resultados pré e pós-teste, Anova com escores de ganho, Ancova<sup>2</sup> e Anova depois da categorização dos sujeitos no pré-teste (*blocking*).

Um fato relevante dessa pesquisa foi o levantamento e a análise de variáveis alternativas (dados secundários) para confrontar com as informações coletadas (dados primários). Para tanto, foram verificados o Produto Interno Bruto, a taxa de juros, a taxa de desemprego, e as mudanças na legislação, os quais indicaram que, na ausência do tratamento, os resultados seriam opostos aos que foram obtidos. Pode-se dizer que os treinamentos realizados pelo governo têm produzido resultados no desempenho dos treinandos, na mudança das instituições e na geração de emprego para as pessoas com deficiência mental.

Em outro tipo de estudo, [Luo e Dappen \(2005\)](#) ressaltaram as avaliações baseadas em objetivos por meio de métodos mistos (*mixed-methods*) em que se combinam métodos quantitativo e qualitativo, capazes de abarcar diversas perspectivas, métodos, dados e valores, os quais podem ser mais eficientes em detectar as fraquezas dos processos de avaliação. Complementando esse posicionamento, [Lion e Martini \(2006\)](#) chamam a atenção para o fato de a quantificação dos *outputs* e dos resultados representarem indubitavelmente o centro da análise avaliativa.

No caso de campanhas de saúde pública, segundo [Yanovitzky et al. \(2005\)](#), muitas avaliações tentam tirar conclusões baseadas em dados observados, no efeito de mensagens audíveis sobre as crenças, atitudes e comportamentos. Para fazer inferências causais, nesse tipo de exemplo, é necessário ajustar os efeitos estimados da campanha para uma possível seleção de vieses em virtude das diferenças sistemáticas

---

<sup>2</sup> Segundo Dancey e Reidy (2006), um dos delineamentos mais comuns em que a Ancova é utilizada é o pré-pós-teste. Nesse caso, o pré-teste é usado como co-variável. Às vezes, deseja-se controlar ou remover o efeito do pré-teste para poder perceber possíveis diferenças devidas à intervenção feita. A utilização da diferença de escores por meio da Anova não permite isso, uma vez que o pré-teste está, geralmente, correlacionado ao resultado (diferença) e assim a variação nos valores do pré-teste não é removida.

entre os respondentes expostos à campanha (grupo experimental) e os não-expostos (grupo controle). Especificamente, é necessário ajustar os efeitos do impacto das variáveis confundidoras<sup>3</sup>, que, provavelmente, são determinantes da exposição e dos resultados da campanha.

[Lion e Martini \(2006\)](#) lembram que o exame dos efeitos causais de intervenções sobre usuários de um programa, em comparação aos resultados que teriam ocorrido sem a intervenção, envolve problemas de estimação.

À semelhança de outros métodos de avaliação com seleção de vieses, tal como o da regressão múltipla e os da variável instrumental (*instrumental variable methods*), a metodologia dos escores propensos (*propensity scores methodology*) oferece uma maneira simples de ajustar as estimativas dos efeitos à exposição de campanhas com a seleção de vieses. O ajuste do escore propenso é descrito pelos referidos autores como um método de dois estágios: o primeiro, envolvendo o equilíbrio das variáveis confundidoras por intermédio dos grupos de tratamento, foi apresentado em cinco passos, e o segundo, relacionado à estimação real dos efeitos de campanha dos ajustados escores propensos, foi discutido no sexto passo.

Os passos descritos pelos autores anteriormente são: (1) seleção de um conjunto de covariáveis (variáveis confundidoras em potencial); (2) determinação do desequilíbrio inicial dos confundidores (teste *t*, para tratamentos de dois níveis e Anova, para tratamentos com mais de dois níveis e, para confundidores dicotômicos, teste da amostra dupla das diferenças em proporções, para tratamentos só com dois níveis e regressão logística, para tratamento com mais de dois níveis); (3) estimação dos escores propensos (análise discriminante, regressão logística); (4) estratificação dos escores propensos (criar cinco estratos baseados nos escores propensos podem remover aproximadamente 90 % do desequilíbrio inicial em cada um dos confundidores dos grupos de tratamento); (5) verificação do equilíbrio dos confundidores

---

<sup>3</sup> Segundo Yanovitzky et al. (2005), confundidoras são aquelas variáveis que podem explicar tanto as variações relativas à exposição à campanha quanto dos resultados, mas elas mesmas não são causadas pela exposição à campanha nem pelos resultados, por exemplo, a idade, o sexo, a raça.

nos grupos de tratamento (métodos de subclassificação, por exemplo, Anova *two-way*) e (6) cálculo da estimativa dos efeitos ajustados dos confundidores.

Para avaliar um programa de fundos relacionado a projetos de R & D, [Sohn et al. \(2007\)](#) propuseram o modelo de equações estruturais (SEM) considerando três aspectos: *output*, resultado (*outcome*) e impacto. Definiram *output* como desempenho tecnológico; resultado, como desempenho do negócio, da gestão e da produção industrial; e impacto, como efeito em longo prazo da melhoria da competitividade nacional e benefícios gerais.

Esses autores usaram o método do quadrado mínimo parcial (PLS) para estimar a SEM e verificar a relação entre os fatores. Enfatizaram, igualmente, que a estimação de máxima verossimilhança (MLE) e o quadrado mínimo parcial (PLS) são ferramentas comuns para estimar a SEM. Embora MLE seja largamente usada, ainda existem limitações, uma vez que ainda necessita não somente da suposição distribucional, mas também de grande número de amostras.

Todas as variáveis foram medidas utilizando-se de uma escala tipo Likert de 5 pontos, o número mais alto representando o melhor desempenho ou satisfação. Metade dos dados da pesquisa foi obtida de visitas às companhias e de entrevistas face a face. O restante dos dados foi obtido da internet. Antes da estimação da SEM proposta, uma análise confirmatória foi conduzida com o objetivo de validar as relações entre as variáveis observadas e as latentes. A confiabilidade do instrumento de pesquisa foi testada pelo alfa de Cronbach, e os resultados mostraram que todos os fatores tiveram um valor maior que 0,70, exceto para o fator *input* R & D.

Como resultado final, [Sohn et al. \(2007\)](#) relatam que o grau com que o modelo de PLS alcança seu objetivo pode ser determinado pelos valores de  $R^2$  dos construtos dependentes (endógenos). Uma das diferenças entre LISREL e o PLS é que, neste último, não existe medida de ajuste geral adequada. De acordo com a literatura, os valores de  $R^2$  para os

construtos endógenos variam de um baixo de 12 % a um alto de 64 %. Para essa pesquisa, foram encontrados valores de  $R^2$  de 17,92 % para o mais baixo e 71,59 % para o mais alto.

Em outro estudo de avaliação de programa, [Jasuja et al. \(2005\)](#) verificaram a relação entre seis características de coligações comunitárias (tamanho da região coberta, estrutura, representação profissional, recursos compartilhados, tamanho, amplitude de prevenção da representação dos *stakeholders*) na predição do progresso organizacional comunitário voltado para a adoção de programas de prevenção contra as drogas. Foram utilizados dados primários de lideranças comunitárias ( $N = 533$ ) de 24 cidades, em um experimento randomizado de múltiplas cidades e estados visando avaliar os efeitos de um treinamento, via televisão, para a adoção de programas de prevenção contra drogas baseados em evidências.

A análise fatorial confirmatória apresentou boa medida de ajuste dos dados do modelo em quatro das referidas características ( $\chi^2 = 6,787$ ;  $gI = 4$ ;  $NFI = 0,993$ ;  $CFI = 0,997$ ). A modelagem de equação estrutural mostrou que essas características tiveram relações positivamente significativas para o progresso organizacional, à exceção do tamanho da região coberta ( $\chi^2 = 5,356$ ;  $gI = 3$ ;  $NFI = 0,994$ ;  $CFI = 0,997$ ). Com base nesses resultados, verifica-se que as coligações que têm uma clara estrutura, foco profissional, recursos compartilhados e o menor serviço regional são provavelmente as que mostram mais progresso na adoção de programas preventivos contra a droga com base em evidências.

Ainda nesse contexto, [Diório e Gomide \(2004\)](#) analisaram o efeito do Programa Bom Aluno (PBA) sobre a escolarização e a profissionalização de 629 jovens. Para isso, estudaram a evolução da escolaridade sem interrupção, a frequência e o tipo de instituição formal de ensino (pública ou privada), o exercício e o tipo de trabalho nos grupos experimental e controle, sendo este último grupo composto de alunos que apresentavam perfil escolar semelhante aos dos estudantes que

ingressaram no PBA com nota mínima de 7,0 em todas as disciplinas, em todos os anos escolares e provenientes de escolas circunvizinhas.

Os autores verificaram efeito positivo do programa no nível superior de ensino ( $\chi^2 = 34,164$ ;  $p < 0,001$ ), no qual, se constatou que havia três vezes mais participantes do grupo controle que do experimental que não freqüentavam a universidade. Foram identificados 62,9 % de alunos do PBA, do nível superior, que trabalhavam e estudavam em comparação com apenas 37,9 % ( $\chi^2 = 8,672$ ;  $p < 0,003$ ) do grupo controle. Esses autores concluíram que as ações do PBA foram efetivas quanto a alcançar, manter o estudo e exercer o trabalho no nível superior de ensino.

Em uma avaliação do Pronaf Infra-estrutura e Serviços Municipais, do Estado de Santa Catarina, [Varaschin et al. \(2002\)](#) verificaram se os objetivos e as estratégias foram bem-sucedidos; se os projetos previstos como meios de concretização dos objetivos foram bem-executados e nas condições previstas (de conteúdo, volume, prazo e custo) e, por fim, se os resultados foram atingidos.

O método empregado para avaliar o desempenho dos projetos focou cinco temáticas: o planejamento, a participação da comunidade, o gerenciamento, o monitoramento e a análise dos resultados. O planejamento foi avaliado segundo três critérios (planejamento estratégico, planejamento operacional e orçamento) e nove indicadores/ itens. O gerenciamento foi analisado por dois critérios (gerenciamento de projetos e de pessoas) e dez indicadores. A participação da comunidade foi verificada por meio de dez itens. Para avaliar o processo de monitoramento, informação e controle, foram utilizados seis indicadores. Por fim, para analisar os resultados, foram empregados 2 critérios (resultados institucionais e indicadores de desenvolvimento social e econômico) e 13 itens.

A amostra foi composta pela equipe técnica e administrativa que atua diretamente na execução dos projetos no município; pelos membros do Conselho Municipal de Desenvolvimento Rural (CMDR) e pelos agricultores beneficiários diretos dos projetos.

Para o levantamento de informações primárias, foram utilizados quatro instrumentos de coleta de dados. O primeiro, aplicado ao grupo executor dos projetos e atividades, foi composto de 38 perguntas e dividido em 3 etapas: na primeira empregou-se uma escala de cinco pontos para medir o grau de eficiência no desempenho da temática tratada (1 = nunca; 2 = poucas vezes; 3 = algumas vezes; 4 = a maioria das vezes e 5 = todas as vezes); na segunda utilizou-se de perguntas que exigiam respostas dicotômicas (sim/não) para expressar se faz ou não, se existe ou não e, na terceira, perguntas voltadas a levantar dados quantitativos conhecidos ou disponíveis no local onde o plano de desenvolvimento rural é gerenciado.

O segundo instrumento aplicado aos membros do CMDR, contendo 23 das 38 perguntas efetuadas aos técnicos, foi conduzido para verificar o grau de afinidade e de interação entre os gestores do plano e o CMDR, bem como para verificar se a percepção dos gestores era condizente com a dos agricultores. O terceiro, aplicado à equipe municipal executora do Plano Municipal de Desenvolvimento Rural, por meio das mesmas 23 perguntas selecionadas anteriormente, serviu para avaliar o grau de execução dos projetos e os resultados obtidos. A última fonte de informação primária foi uma visita aos projetos financiados para verificar junto aos agricultores o andamento dos processos organizativos e da gestão dos projetos, por meio de conversas informais.

Para a coleta de informações secundárias, as principais fontes de informações foram o Plano Municipal de Desenvolvimento Rural; os projetos apoiados pelo Pronaf Infra-estrutura e os relatórios de acompanhamento desses projetos em Santa Catarina.

No que se refere aos resultados encontrados por [Varaschin et al. \(2002\)](#), destacam-se os projetos de agregação de valor (49 %) como insatisfatórios ou regulares. Esses projetos incluem a construção de pequenas agroindústrias, unidades de beneficiamento de produtos e unidades de comercialização de produtos agrícolas. Por sua vez, o grupo de projetos ligados à aquisição de máquinas e de equipamentos agrícolas obteve um grau de satisfação (de bom a ótimo) em 86 % dos

casos. Isso vai ao encontro da opinião dos conselheiros municipais, que consideram esse grupo de projetos muito importante para os agricultores.

Por fim, os referidos autores ponderam que o Pronaf Infra-estrutura, além de ser fonte de recursos para o financiamento de projetos, contribuiu para a melhoria dos serviços oferecidos aos agricultores familiares e introduziu uma nova fórmula de atuação do serviço público estadual e federal junto aos pequenos municípios, por meio do apoio financeiro e técnico aos projetos de agricultura familiar.

Com base nesse referencial teórico, pode-se considerar que a avaliação de programa é uma espécie de ferramenta de trabalho aplicada ao campo de pesquisa multidisciplinar, envolvendo conceitos, teorias e métodos específicos de operacionalização. É vista como importante mecanismo de apoio para a tomada de decisões de gestores de programas. Sua utilização, tendo como base os parâmetros da Psicologia Social, como as crenças, as atitudes, os valores, as opiniões e os comportamentos humanos, pode contribuir de forma significativa para o progresso da sociedade.

## Referências

ALBUQUERQUE, F. J. B. de; MASCAREÑO, R. P. Considerações não-ortodoxas sobre as cooperativas e o cooperativismo. **Psicologia e Sociedade**, v. 11, n. 2, p. 23-41, 1999.

AMORIM, F. R. S. **O Banco do Nordeste e o Programa Nacional de Fortalecimento da Agricultura Familiar – Pronaf**. Fortaleza: Banco do Nordeste, 2003. Cartilha.

ARRETICHE, M. T. S. Tendências no estudo sobre avaliação. In: RICO, E. M. (Org.). **Avaliação de políticas sociais: uma questão em debate**. 3. ed. São Paulo: Cortez, 2001. p. 29-49.

AZEVEDO, C. M. de; SHIKIDA, P. F. A. Assimetria de informação e o crédito agropecuário: p caso dos cooperados da Coamo-Toledo (Pr). **Revista de Economia e Sociologia Rural**, v. 42, n. 2, p. 267-292, 2004.

BANCO CENTRAL DO BRASIL. **Crédito rural: Programa Nacional de Fortalecimento da Agricultura Familiar (Pronaf)** - 10. Disponível em: <<http://www4.bcb.gov.br/NXT/>

gateway.dll?f=templates&fn=default.htm&vid=nmsDenorMCR:idvDenorMCR>. Acesso em: 17 maio 2007.

BELLONI, I.; MAGALHÃES, H. de; SOUSA, L. C. de. **Metodologia de avaliação em políticas públicas**. 2. ed. São Paulo: Cortez, 2001. (Coleção Questões de Nossa Época, 75).

BERGER, P. L. **A construção social da realidade**: tratado de sociologia do conhecimento. Petrópolis: Vozes, 1973.

BRASIL. Ministério do Desenvolvimento Agrário. **Portal do Ministério do Desenvolvimento Agrário**: resultados da pesquisa. Brasília, 2005. Disponível em: <[www.mda.gov.br/index.php?dmnid=29&scoid=&scant=](http://www.mda.gov.br/index.php?dmnid=29&scoid=&scant=)>. Acesso em: 7 out. 2005.

BRASIL. Ministério do Desenvolvimento Agrário. **Apresentação – crédito rural**. Disponível em: <<http://www.mda.gov.br/saf/index.php?scoid=813>>. Acesso em: 16 maio 2007a.

BRASIL. Ministério do Desenvolvimento Agrário. **Resultados da pesquisa**. Disponível em: <<http://www.mda.gov.br/index.php?dmnid=29&scoid=&scant=>>. Acesso em: 16 maio 2007b.

BRASIL. Tribunal de Contas da União. Secretaria de Fiscalização e Avaliação de Programas de Governo. **Técnica de auditoria**: indicadores de desempenho e mapa de produtos. Brasília, 2000.

BRASIL. Tribunal de Contas da União. Secretaria de Fiscalização e Avaliação de Programas de Governo. **Técnicas de auditoria**: marco lógico. Brasília, 2001.

BRASIL. Tribunal de Contas da União. Secretaria de Fiscalização e Avaliação de Programas de Governo. **Técnicas de auditoria**: análise stakeholder. Brasília, 2002.

CAMPBELL, D. T.; STANLEY, J. C. **Delineamentos experimentais e quase-experimentais de pesquisa**. São Paulo: EPU: Editora da Universidade de São Paulo, 1979.

CAMPOS, L. F. de L. **Métodos e técnicas de pesquisa em psicologia**. Campinas: Alínea, 2001.

CANO, I. **Introdução à avaliação de programas sociais**. 2. ed. Rio de Janeiro: FGV, 2004. (Coleção FGV Prática).

CASTRO, G. A. de O. **Um modelo integrado de avaliação do impacto do treinamento no trabalho – IMPACT**. 1999. Tese (Doutorado) - Instituto de Psicologia, Universidade de Brasília, Brasília, DF.

CONE, J. D. **Evaluating outcomes**: empirical tools for effective practice. Washington: American Psychological Association, 2001.

COZBY, P. C. **Métodos de pesquisa em ciências do comportamento**. São Paulo: Atlas, 2003.

CRONBACH, L. J. **Designing evaluations of educacional and social programs**. San Francisco: Jossey-Bass, 1982.

DANCEY, C. P.; REIDY, J. **Estatística sem matemática para psicologia: usando SPSS para windows**. 3. ed. Porto Alegre: Artmed, 2006.

DIÓRIO, Z. M.; GOMIDE, P. I. C. Ascensão escolar e profissionalização de bons alunos de baixa renda: avaliação de um programa brasileiro. **Psicologia: reflexão e crítica**, v. 17, n 3, p. 359-366, 2004.

DONALDSON, S. I.; GOOLER, L. E. Theory-driven evaluation in action: lessons from a \$20 million statewide work and health initiative. **Evaluation and Program Planning**, v. 26, p. 355-366, 2003.

DUGARD, P.; TODMAN, J. Analysis of pre-test-pos-test control group designs in educational research. **Educational Psychology**, v. 15, n. 2, p. 181-198, 1995.

FARIA, C. A. P. de. A política da avaliação de políticas públicas. **Revista Brasileira de Ciências Sociais**, v. 20, n. 59, p. 97-109, 2005.

FERREIRA, C. R. C. **Como o agricultor familiar pode conseguir e manter o financiamento rural e como se dá a relação com os bancos**. Disponível em: <<http://www.mda.gov.br/saf/arquivos/1137912740.doc>>. Acesso em: 16 maio 2007a.

FERREIRA, C. R. C. **Como negociar e assegurar as garantias que serão dadas ao banco quando se necessita de um financiamento rural do Pronaf**. Disponível em: <<http://www.mda.gov.br/saf/arquivos/1137912741.doc>>. Acesso em: 16 maio 2007b.

IBGE. **Censo agropecuário de 1995-1996**. Disponível em: <[http://www.ibge.gov.br/home/estatistica/economia/agropecuaria/censoagro/1995\\_1996/brasil/tabela4brasil.shtm](http://www.ibge.gov.br/home/estatistica/economia/agropecuaria/censoagro/1995_1996/brasil/tabela4brasil.shtm)>. Acesso em: 28 ago. 2007.

JASUJA, G. K.; CHOU, C. P.; BERNSTEIN, K.; WANG, E.; MCCLURE, M.; PENTZ, M. A. Using structural characteristics of community coalitions to predict progress in adopting evidence-based prevention programs. **Evaluation and Program Planning**, v. 28, p. 173-184, 2005.

LEE, Y. F.; ALTSCHULD, J. W.; WHITE, J. L. Effects of multiple stakeholders in identifying and interpreting perceived needs. **Evaluation and Program Planning**, v. 30, p. 1-9, 2007.

LION, C.; MARTINI, P. The evaluation of a complex social program: lessons learned from the experience of the European Social Fund. **Evaluation and Program Planning**, v. 29, p. 1-9, 2006.

LUO, M.; DAPPEN, L. Mixed-methods design for an objective-based evaluation of a magnet school assistance project. **Evaluation and Program Planning**, v. 28, p. 109-118, 2005.

MARINHO, A.; FAÇANHA, L. O. **Programas sociais: efetividade, eficiência e eficácia** como dimensões operacionais da avaliação. Rio de Janeiro: IPEA, 2001. (Texto para Discussão, 787).

MOURÃO, L. **Avaliação de programas públicos de treinamento: um estudo sobre o impacto no trabalho e na geração de emprego**. 2004. Tese (Doutorado) - Instituto de Psicologia, Universidade de Brasília, Brasília, DF.

POSAVAC, E. J.; CAREY, R. G. **Program evaluation: methods and case studies**. 6. ed. New Jersey: Prentice Hall, 2003.

RENGER, R.; HURLEY, C. From theory to practice: lessons learned in the application of the ATM approach to developing logic models. **Evaluation and Program Planning**, v. 29, p. 106-119, 2006.

RIBEIRO, R. P.; RIBEIRO, O. C.; VASCONCELOS, J. R. P.; GUIMARÃES, P. M. **Projetos de captação de recursos para organizações de P&D: conceitos, metodologias e informações básicas**. Brasília: Embrapa-SSE, 1998.

SCHALOCK, R. L.; BONHAM, G. S. Measuring outcomes and managing for results. **Evaluation and Program Planning**, v. 26, p. 229-235, 2003.

SCRIVEN, M. Types of evaluation and types of evaluator. **Evaluation Practice**, v. 17, n. 2, 151-161, 1996.

SCRIVEN, M. Minimalist theory: the least theory that practice requires. **American Journal of Evaluation**, v. 19, n. 1, p. 57-70, 1998.

SCRIVEN, M. Evaluation: future tense. **American Journal of Evaluation**, v. 22, n. 3, p. 301-307, 2001.

SILVA, P. L. B. (Coord.). **Modelo de avaliação de programas sociais prioritários: relatório final**. Campinas: Unicamp, 1999.

SILVA, M. O. da S. e. Avaliação de políticas e programas sociais e metodológicos. In SILVA, M. O. da S. e. (Org.). **Avaliação de políticas e programas sociais: teoria e prática**. São Paulo: Veras, 2001. p. 37-93.

SOHN, S. Y.; JOO, Y. G.; HAN, H. K. Structural equation model for the evaluation of national funding on R & D project of SMEs in consideration with MBNQA criteria. **Evaluation and Program Planning**, v. 30, p. 10-20, 2007.

SRIDHARAN, S.; GO, S.; ZINZOW, H.; GRAY, A.; BARRET, M. G. Analysis of strategic plans to assess planning for sustainability of comprehensive community initiatives.

**Evaluation and Program Planning**, v. 30, p. 105-113, 2007.

UNRUH, D. Using primary and secondary stakeholders to define facility-to-community transition needs for adjudicated youth with disabilities. **Evaluation and Program Planning**, v. 28, p. 413-422, 2005.

VARASCHIN, V. M.; HEIDEN, F. C.; PAUL, J. M.; SILVA, J. C. M. **Avaliação do Pronaf-estrutura e serviços municipais**. Florianópolis: Instituto de Planejamento e Economia Agrícola de Santa Catarina, 2002.

VOLPI, R.; BRESSAN, M. Bases para programas de qualificação e habilitação de mão-de-obra para o agronegócio do leite. In: GOMES, A. T.; LEITE, J. L. B.; CARNEIRO, A. V. (Org.). **O agronegócio do leite no Brasil**. Juiz de Fora: Embrapa Gado de Leite, 2001. p. 111-122.

YANOVITZKY, I.; ZANUTTO, E.; HORNIK, R. Estimating causal effects of public health education campaigns using propensity score methodology. **Evaluation and Program Planning**, v. 28, p. 209-220, 2005.

## Referências recomendadas

Considerando que esta publicação foi elaborada com base na tese do primeiro autor, para maior aprofundamento deste estudo, recomenda-se a leitura dos seguintes trabalhos extraídos da mesma fonte:

ROCHA, F. E. de C.; ALBUQUERQUE, F. J. B. de; MARCELINO, M. Q. dos S.; DIAS, M. R.; PINHEIRO, J. de Q. **Aplicação da análise de conteúdo na perspectiva de Bardin em uma aproximação avaliativa do Pronaf-PB**. Planaltina, DF: Embrapa Cerrados, 2008. (Embrapa Cerrados. Boletim de Pesquisa e Desenvolvimento, 201).

ROCHA, F. E. de C.; ALBUQUERQUE, F. J. B. de; DIAS, M. R.; COELHO, J. A. P. de M.; MARCELINO, M. Q. dos S. **Intenção de pagamento do crédito (Pronaf): uma aplicação da Teoria da Ação Planejada**. Planaltina, DF: Embrapa Cerrados, 2008. (Embrapa Cerrados. Boletim de Pesquisa e Desenvolvimento, 202).

ROCHA, F. E. de C.; ALBUQUERQUE, F. J. B. de; TRÓCCOLI, B. T.; COELHO, J. A. P. de M.; MARCELINO, M. R. **Modelo de avaliação de programa: uma referência para o monitoramento do Pronaf na Paraíba**. Planaltina, DF: Embrapa Cerrados, 2008. (Embrapa Cerrados. Boletim de Pesquisa e Desenvolvimento, 203).

ROCHA, F. E. de C.; ALBUQUERQUE, F. J. B. de; PINHEIRO, J. de Q.; COELHO, J. A. P. de M.; MARCELINO, M. R. **O Pronaf na Paraíba: uma avaliação quantitativa com enfoque psicossocial**. Planaltina, DF: Embrapa Cerrados, 2008. (Embrapa Cerrados. Boletim de Pesquisa e Desenvolvimento, 204).

ROCHA, F. E. de C.; ALBUQUERQUE, F. J. B. de; COELHO, J. A. P. de M.; MARCELINO, M. Q. dos S. **Elaboração de instrumento de avaliação do Pronaf para agricultores com baixa escolaridade**. Planaltina, DF: Embrapa Cerrados, 2008. (Embrapa Cerrados. Documentos, 210).

ROCHA, F. E. de C.; ALBUQUERQUE, F. J. B. de; MARCELINO, M. Q. dos S.; DIAS, M. R. **Aplicações da Teoria da Ação Planejada: uma possibilidade para estudos comportamentais com agricultores familiares**. Planaltina, DF: Embrapa Cerrados, 2008. (Embrapa Cerrados. Documentos, 212).

## Evaluation programs: theoretical and methodological aspects in the small holder context

---

**Abstract** – *This study aims to present theoretical information from literature about evaluation of social programs, which were used later on to evaluate the Pronaf (a credit line for small farmers implemented by the Brazilian Federal Government) in Paraíba State. Several approaches about the evaluation logic models were identified, in which data and information are gathered through tables or input lists, activities, products, results and impacts. Other strategy is the evaluation focused on results, in which was verified the recommendation of quasi-experimentation design, with control group, to isolate results of the specific program. Moreover, concepts related to hard indexes and perception data were also presented to facilitate visualization of how determined variable criterion occurs in the real world, i. e., how a perception data can be confirmed or not by the hard index. Finally, efficiency and effectiveness matters were commented.*

*Index terms: social psychology, program evaluation, small holders, Pronaf.*