

✓ Avaliação de Programas Sociais em uma perspectiva sistêmica, plural e progressista: conceitos, tipologias e etapas¹

Paulo de Martino Jannuzzi²

¹ Este texto atualiza reflexões já expostas anteriormente em outras publicações, em especial, em Jannuzzi (2016a).

² Paulo Jannuzzi é professor da Escola Nacional de Ciências Estatísticas do IBGE, em estágio-posdoutoral na EBAPE/FGV. Professor da Faculdade Cesgranrio. Foi Secretário de Avaliação e Gestão da Informação do Ministério de Desenvolvimento Social e Combate à Fome (2011-2016).

Resumo: A Avaliação de políticas e programas desenvolveu-se historicamente como campo tecnocientífico vinculado ao ritmo e demandas de informação qualificada para Políticas Públicas. Na América Latina tal desenvolvimento é bem mais recente que no contexto de países centrais, tendo ocorrido nos últimos 30 a 40 anos. Mais recentemente, nos últimos 15 anos, visões mais multidisciplinares (na interpretação das problemáticas sociais), ecléticas (em metodologias de investigação) e plurais (em valores e critérios avaliativos) têm sido propostas na condução de Avaliações de Políticas, programas e projetos na região. Este texto procura contribuir para fortalecimento desse campo tecnopolítico contra-hegemônico de Avaliação, ao sistematizar conceitos e técnicas a partir da experiência prática de conduzir avaliações orientadas a subsidiar diretamente à gestão de políticas e programas públicos no Brasil. Em diferentes seções, são apresentados conceitos, tipologias de Avaliação e as várias etapas para especificação de um estudo avaliativo.

Palavras-chave: Avaliação de Programas; Técnicas de avaliação; Etapas de pesquisa

EN Evaluation of Social Programs in a systemic, plural and progressive perspective: concepts, typologies and stages

Abstract: The evaluation of policies and programs has historically developed as a techno-scientific field linked to the pace and demands of qualified information for Public Policies. In Latin America, this development is much more recent than in the context of central countries, having occurred in the last 30 to 40 years. More recently, in the past 15 years, there have been proposed more multidisciplinary (in the interpretation of social issues), eclectic (in research methodologies) and plural (in values and evaluation criteria) views in conducting Policy Assessments, programs and projects in the region. This text seeks to contribute to the strengthening of this counter-hegemonic technopolitical field of Evaluation, by systematizing concepts and techniques from the practical experience of conducting assessments aimed at subsidizing directly the management of public policies and programs in Brazil. In different sections, concepts, types of evaluation and the various steps for specifying an evaluative study are presented.

Key-words: Program Evaluation; Evaluation techniques; Research stages

ES Evaluación de Programas Sociales en una perspectiva sistémica, plural y progresista: conceptos, tipologías y etapas

Resumen: La evaluación de políticas y programas se ha desarrollado históricamente como un campo tecnocientífico vinculado al ritmo y las demandas de información calificada para Políticas Públicas. En América Latina, este desarrollo es mucho más reciente que en el contexto de los países centrales, habiendo ocurrido en los últimos 30 a 40 años. Más recientemente, en los últimos 15 años, se han propuesto enfoques más multidisciplinares (en la interpretación de los problemas sociales), eclécticas (en metodologías de investigación) y plurales (en valores y criterios de evaluación) al realizar evaluaciones de políticas, programas y proyectos en la región. Este texto busca contribuir al fortalecimiento de este campo tecnopolítico contrahegemónico de evaluación, sistematizando conceptos y técnicas a partir de la experiencia práctica de realizar evaluaciones dirigidas a subsidiar directamente la gestión de políticas y programas públicos en Brasil. En diferentes secciones, se presentan conceptos, tipos de evaluación y los diversos pasos para especificar un estudio evaluativo.

Palabras-clave: Evaluación del programa; Técnicas de evaluación; Pasos de investigación

FR Évaluation des programmes sociaux dans une perspective systémique, plurielle et progressiste: concepts, typologies et étapes

Résumé: L'évaluation des politiques et des programmes s'est historiquement développée comme un domaine technoscientifique lié au rythme et aux demandes d'informations qualifiées pour les politiques publiques. En Amérique latine, cette évolution est beaucoup plus récente que dans le contexte des pays centraux, survenue au cours des 30 à 40 dernières années. Plus récemment, au cours des 15 dernières années, des vues plus multidisciplinaires (dans l'interprétation des questions sociales), éclectiques (dans les méthodologies de recherche) et plurielles (dans les valeurs et les critères d'évaluation) ont été proposées dans la conduite des évaluations des politiques, des programmes et des projets dans la région. Ce texte cherche à contribuer au renforcement de ce champ technopolitique contre-hégémonique de l'évaluation, en systématisant les concepts et les techniques issus de l'expérience pratique de la conduite d'évaluations visant à subventionner directement la gestion des politiques et programmes publics au Brésil. Dans différentes sections, les concepts, les types d'évaluation et les différentes étapes de spécification d'une étude évaluative sont présentés.

Mots-clés: évaluation de programme; Techniques d'évaluation; Étapes de recherche

Introdução

Em uma perspectiva histórica, a Avaliação de políticas e programas desenvolveu-se como campo tecnocientífico vinculado ao ritmo e demandas da estruturação do Estado de Bem-Estar Social. Quanto mais as Políticas Públicas foram sendo pautadas e implementadas por governos eleitos para responder às demandas da sociedade, mais informação qualificada foi necessária para subsidiar as atividades de gestão. Se na Europa central esse processo de estruturação – das Políticas Públicas, Avaliação e Sistema Estatístico – se iniciou no começo do século XX, na América Latina – e também em Portugal e Espanha – ele é bem mais recente, tendo ocorrido nos últimos 30 a 40 anos. Afinal, é nesse período que se dá a redemocratização em diversos países na região, viabilizando a eleição de governos com agendas mais progressistas, o fortalecimento de movimentos sociais e, assim, a institucionalização de marcos normativos mais amplos em termos de direitos sociais³.

Políticas públicas para acesso universal à educação básica e saúde, garantia da segurança alimentar, provimento de renda mínima para população mais vulnerável, regulação ambiental e ações afirmativas e compensatórias para segmentos populacionais têm sido implantados em diversos países ibero-americanos nas últimas décadas, impondo a elaboração de avaliações diagnósticas, de implementação e de resultados e impactos. A profissionalização técnica dos serviços públicos nos vários âmbitos de governo na região também é outro fator importante a destacar para o aumento da demanda de informações mais complexas sobre Políticas Públicas, refletindo-se sobre a agenda de trabalho de institutos nacionais de estatísticas, centros de pesquisas e também nos ministérios setoriais (JANNUZZI 2020).

Se o contexto de redemocratização favoreceu o alargamento da pauta de Políticas Públicas e, por conseguinte, da demanda por Avaliação, o *zeitgeist* liberal que perseguiu a Administração Pública na região em boa parte desse período acabou por conformar o campo com forte viés técnico-formal-positivista e orientado por critérios economicistas. Como bem registrou Gussi (2019, p.171-172):

Esse contexto leva à perspectiva avaliativa fomentada pelos organismos internacionais, como o Banco Mundial e BID, e suas políticas de financiamento,

que remetem a uma avaliação nos marcos políticos neoliberais. No Brasil, essa perspectiva tomou forma com a Reforma de Estado, nos anos 1990, quando fortaleceu-se um modelo de avaliação de gestão pública ajustado às demandas do Banco Mundial ...

A essa agenda política circunscreve-se um modelo de avaliação de programas, projetos e políticas, sobretudo governamentais, baseados em critérios pré-definidos de eficiência, eficácia e efetividade.

Quase sempre reduzidas à dimensão econômica, essas avaliações têm por intuito demonstrar o sucesso ou fracasso das políticas a partir da construção de indicadores, notadamente quantitativos, que revelam a otimização da relação custo-benefício, previamente calculada, em relação ao investimento realizado na execução de programas, projetos e políticas...

Trata-se, antes, de imputar um modelo de avaliação de cunho técnico-formal, entendido como um modelo único e universal de avaliação, que traz em si uma proposição epistemológica funcional e positivista, desconsiderando os sujeitos sociais envolvidos nas políticas, bem como os contextos sócio-políticos e culturais nacionais, regionais e locais onde essas políticas realizam-se, e as contradições neles inerentes.

De fato, o protagonismo de organizações multilaterais de fomento na disseminação do pacote tecnocrático da Nova Gestão Pública nos anos 1980-1990, embalado sob os rótulos de “Gestão-por-Resultados” e “Política-Pública-orientado-por evidências”, acabou influenciando os rumos da institucionalização e as práticas de avaliação no Brasil e outros países na região. Manuais e livros no campo da Avaliação, produzidos em particular nos EUA, formaram cabeças e mentes de parte expressiva da comunidade epistêmica e de práticas dos países ibero-americanos, sobretudo nas escolas de economia e áreas de finanças e controle da Administração Pública. Mas, modelos e práticas avaliativas assentadas nos pressupostos

³ Vale registrar que Portugal e Espanha vivenciaram governos autoritários – de Salazar/Marcelio Caetano e Francisco Franco, respectivamente – até anos 1970.

tecnocráticos da eficiência e custo-benefício, preconizado pela comunidade epistêmica dos *econo-avaliadores*, ou da conformidade burocrática, entronizada pelos *audito-avaliadores*, não parecem estar conduzindo a Políticas Públicas mais efetivas no Brasil e outros países. Ao contrário, têm levado à descontinuidade de políticas e programas meritórios que não tiveram tempo para serem implantados adequadamente e à baixa inovação no desenho de intervenções necessárias aos problemas vivenciados pela população (JANNUZZI 2016b).

Mais recentemente, nos últimos 15 anos, visões mais multidisciplinares (na interpretação das problemáticas sociais), ecléticas (em metodologias de investigação) e plurais (em valores e critérios avaliativos) têm sido propostas na condução de Avaliações de Políticas, programas e projetos na região (MINAYO *et al.* 2005; PAES-SOUSA,VAITSMANN 2011;GUSSI;OLIVEIRA 2017; GUSSI 2019, SAGI 2016). Conceitos, técnicas, práticas e abordagens mais circunstanciadas de Avaliação tem sido propostas, documentadas e disseminadas para responder ao contexto complexo das problemáticas sociais vivenciados na região, de operação dos programas em um quadro de heterogeneidade de capacidade de gestão em nível subnacional e frente à fragilidade do marco institucional dos Direitos e Políticas no país.

Na perspectiva de contribuir para fortalecimento desse campo tecnopolítico contra-hegemônico de Avaliação, esse texto sistematiza conceitos e técnicas a partir da experiência prática de conduzir avaliações orientadas a subsidiar à gestão de políticas e programas públicos no Brasil segundo uma perspectiva mais republicana e não economicista. Em diferentes seções, apresenta-se conceitos, tipologias de Avaliação e as várias etapas para especificação de um estudo avaliativo. Trata-se antes de um texto com propósitos didáticos e documentais do que analítico ou ensaístico no campo.

Uma definição pragmática de Avaliação

Há muitas definições de Avaliação na literatura especializada, desde aquelas que assumem a Avaliação em perspectiva mais geral (PATTON, 1997;WORTHERN *et al.*, 2004; ROSSI *et al.*, 2004); como instrumento de sistemas de monitoramento e avaliação

de programas governamentais (MACDAVID; HAWTHORNE, 2006; IMAS,HIST 2009); ou, em uma concepção mais restrita (GERTLER, 2015), de entender Avaliação como um tipo particular de investigação empírica acerca de programas e projetos sociais, como as avaliações de impacto experimental ou *quasi-experimental*

Em uma definição mais pragmática e aplicada, Avaliação refere-se ao conjunto de procedimentos técnicos para produzir informação e conhecimento, em perspectiva interdisciplinar, para desenho *ex-ante*, implementação e validação *ex-post* de programas e projetos sociais, por meio das diferentes abordagens metodológicas da pesquisa social, com a finalidade de garantir o cumprimento dos objetivos de programas e projetos (eficácia), seus impactos mais abrangentes em outras dimensões sociais, ou seja, para além dos públicos-alvo atendidos (efetividade) e a custos condizentes com a escala e complexidade da intervenção (eficiência).

Na definição aqui advogada, avaliação tem o objetivo de produzir evidências, compilar dados e sistematizar estudos que contribuam para o aperfeiçoamento de programas e projetos sociais, além da consecução de seus objetivos. Em tal definição, avaliação não é entendida tão somente como uma investigação com métodos validados cientificamente para analisar diferentes aspectos *sobre* um programa – o que se constituiria em uma investigação de cunho acadêmico –, mas um levantamento consistente, sistemático e replicável de dados, informações e conhecimentos *para* aprimoramento da intervenção programática, versando sobre:

- as características essenciais do contexto de atuação dos programas;
- os públicos-alvo das intervenções;
- o desenho, os arranjos e desafios de implementação;
- os resultados de curto prazo;
- os impactos sociais e de mais longo prazo de um programa;
- os custos operacionais de diferentes formas de intervenção pública.

A definição de avaliação de programas aqui proposta também remete à necessidade de abordagem interdisciplinar na produção de informação e conhecimento sobre os problemas investigados. Por um lado, deve-se lembrar que pobreza, desigualdade, desempenho escolar e mortalidade infantil são questões sociais multideterminadas, isto é, determinadas e influenciadas por uma série de dimensões sociais mais amplas e estruturais. De outro lado, programas públicos são empreendimentos complexos,

que envolvem a contratação de pessoal técnico; disponibilidade de instrumentos; adequação de equipamentos públicos; alocação de recursos monetários; promoção de capacitação, de forma coordenada no tempo e no território. Mapear, pois, as dimensões sociais mais relevantes à formulação de um programa ou os problemas operacionais do mesmo requer uma equipe de avaliadores com diferentes formações acadêmicas e profissionais. Certamente se requer especialistas acadêmicos nas temáticas investigadas, mas não se pode deixar de ter nas equipes de avaliação profissionais que já tiveram experiência na gestão de programas correlatos.

Essa conceituação da área também denota um sentido muito mais abrangente do que entender Avaliação como apreciação somente de resultados e impactos de uma intervenção, para prestação de contas ou alocação orçamentária. Avaliação é, como defende Rossi *et al* (2004), um campo aplicado de produção de conhecimento para subsidiar a gestão e aprimoramento de uma política, programa ou projeto, do seu desenho à análise de mérito. Tampouco, com tal definição, Avaliação pode ser resumida ao que comunidades epistêmicas específicas – randomistas (RAVALLION2009) – têm denominado de Avaliação de Impacto. Experimentos aleatorizados e pesquisas *quasi*-experimentais podem ter aplicações relevantes na avaliação de programas públicos, quando os imperativos éticos, políticos e operacionais não limitam sua utilização, e desde que seus resultados sejam interpretados à luz das limitações concretas com que são realizados (JANNUZZI 2018).

Avaliação Sistêmica parece ser, pois, um termo oportuno e adequado para conferir esse sentido mais amplo do campo como aqui advogado, muito além de apreciações de mérito, com técnicas específicas, ao final de um ciclo de implementação de políticas ou programas.

Tipologias instrumentais de Avaliação

Na bibliografia do campo há inúmeras propostas de classificação das pesquisas e estudos de avaliação, como as sistematizadas em Worthern *et al*. (2004), Imas e Rist (2009). Muitas dessas tipologias têm usos muito particulares em campos temáticos específicos de avaliação, mas com utilidade instrumental limitada para análise

mais geral de programas públicos. Na perspectiva mais pragmática desse texto, há algumas poucas classificações mais essenciais para diferenciar as Avaliações.

A mais básica e simples é a categorização segundo o momento de sua realização em relação ao programa: avaliação *ex-ante*, realizada antes do início de um programa, para caracterização do público-alvo e dimensionamento da problemática social que será o foco da intervenção; e avaliação *ex-post*, para análise comparativa das mudanças ocorridas junto ao público-alvo e aferição de resultados potenciais do programa. Adicionalmente nessa tipologia haveria a Avaliação de processo, voltada ao exame de rotinas operacionais do programa, para identificação de gargalos e problemas de implementação (MINAYO *et al* 2005).

Outra categorização clássica de avaliações é segundo agente avaliador do programa: *avaliação externa*, *avaliação interna*, *mista* e *participativa* (IMAS;RIST 2009). Em tese, avaliações com equipes externas podem garantir maior independência técnica quanto à condução do estudo avaliativo do programa que as realizadas internamente. Equipes internas detêm, em geral, conhecimento sobre os problemas e as dificuldades mais prementes que afetam os programas públicos, mas precisam de apoio técnico de consultores especializados e pesquisas de campo para entender suas causas e estratégias de resolução. Com equipes mistas, garante-se a sinergia entre conhecimento interno de gestão e especialidade técnica externa, gerando produtos com maior potencialidade de apropriação, não apenas pela adequação às demandas mais relevantes como também pela legitimidade conferida pelo envolvimento inicial dos agentes internos ao programa .

Em algumas situações, menos frequentes na avaliação de programas públicos, desenvolve-se abordagens participativas – ou “*quasi*-participativas”-, quarto tipo de avaliação segundo agente avaliador, com participação direta ou com algum nível de interação participativa do público-alvo no processo de coleta e sistematização de informação. Esse tipo de avaliação é empregado quando se procura engajar os beneficiários e usuários na concretização dos objetivos do programa. É como uma pesquisa-ação, em que a obtenção de informações é tão importante quanto promover a busca de soluções para as dificuldades do programa, como em projetos de economia solidária, de cooperativismo e auto-construção de moradias. Em

algumas situações é a única estratégia de coleta, frente as dificuldades de abordagem de problemáticas sociais complexas ou de públicos refratários às estratégias mais convencionais, como as realizadas com população em situação de rua, com jovens em situação de drogadição, com comunidades quilombolas isoladas, entre outras. Naturalmente, nesse tipo de avaliação os parâmetros clássicos de distanciamento do objeto, imparcialidade dos sujeitos e replicabilidade dos resultados, que orientam as metodologias convencionais cedem lugar ao potencial efeito engajador da abordagem participativa na busca de soluções para os problemas identificados.

Cada uma dessas modalidades tem vantagens e limitações. A avaliação externa parece permitir um olhar objetivo não enviesado e não comprometido com a continuidade do programa, conferindo maior isenção à análise, supondo que os consultores são absolutamente profissionais, não preocupados com as chances de contratação em uma a próxima demanda de avaliação. Pelo envolvimento das equipes, a avaliação interna, a mista e a participativa parecem permitir aportes mais circunstanciados na análise dos processos e resultados, além de facilitar a incorporação de melhorias na gestão do programa. Mas podem ser cegas aos problemas que coloquem em xeque a continuidade do programa. Avaliações internas podem ser rápidas e baratas, ao contrário das outras modalidades. Avaliações participativas podem viabilizar aprimoramentos mais efetivos, pelo envolvimento dos beneficiários no sucesso do programa.

Uma tipologia de pesquisas avaliativas menos conhecida, mas particularmente interessante para organizar a avaliação de programas sociais é a proposta por Rossi *et al.* (2004), de classificação de avaliação segundo necessidades de informação para o estágio em se encontra o programa no seu ciclo de vida. São cinco os tipos de avaliação. Para dimensionamento de demandas coletivas, o autor propõe a *avaliação de necessidades ou demanda*, apoiada em fontes de dados já existentes, como os produzidos pelos institutos nacionais de estatística e outros centros oficiais de estudos, além dos cadastros públicos dos ministérios setoriais. Com isso, pode-se garantir um rápido dimensionamento e caracterização da questão social a ser objeto de intervenção pública.

Para a formulação de programa ou projeto para mitigação ou equacionamento da problemática social identificada, em geral, são necessários novos esforços de levantamentos de campo – para aprofundamento do diagnóstico das condições de vida; contexto

econômico; restrições ambientais; capacidade de gestão e oferta de serviços – e de compilação de estudos já realizados na temática, abordando determinantes da problemática em questão e eventuais programas e projetos já idealizados, atividades que constituem o que se denomina *avaliação de desenho*.

Definidos os públicos que serão atendidos e os arranjos operacionais do programa ou projeto social, é preciso colocá-lo em ação, realizando as atividades planejadas, acompanhando a execução dessas mediante indicadores de gestão e de monitoramento, além de identificar problemas na oferta, regularidade e qualidade dos serviços por meio de pesquisas de *avaliação da implementação*. Nesse tipo de avaliação, realizado em geral por meio de estratégias qualitativas ou indicadores provenientes de informação dos sistemas de gestão, procura-se analisar os nós críticos que dificultam a plena operação dos programas como os diferentes aspectos relacionados à disponibilidade de recursos humanos, recursos orçamentários, aos equipamentos públicos, ao relacionamento com atores de outras esferas de governo e agentes envolvidos na operação etc. Esse tipo de avaliação costuma ser denominado também de avaliação formativa.

Reconhecidos e, tanto quanto possível, sanados os desafios da implementação, as demandas de informação e conhecimento voltam-se à *avaliação de resultados e impactos* do programa ou projeto social. Trata-se do momento para uma investigação mais exaustiva sobre os diversos componentes de uma intervenção, abordando não apenas o cumprimento de seus objetivos, mas também seu desenho, arranjos operacionais, impactos sociais mais abrangentes – no tempo e no território –, e sobre a capacidade de inovação e redesenho frente ao contexto dinâmico em que operam os programas e projetos. Cabe nesse momento avaliar se a intervenção programática formulada conseguiu provocar mudanças na realidade social que a originou, considerando naturalmente a complexidade de seu desenho e dos arranjos operacionais, além da criticidade da questão social enfrentada. Esse tipo de avaliação é conhecido também por avaliação somativa.

Identificar o momento adequado para avaliações dessa natureza é um misto de técnica, política e arte, pois avaliações precoces podem colocar a perder a legitimidade de um programa e projeto meritório que ainda não teve tempo de se estruturar. Avaliações tardias podem comprometer recursos e esforços que poderiam ser utilizados de

forma mais eficiente e eficaz na mitigação da problemática social em questão. Em síntese, avaliações são, como toda atividade na gestão pública, empreendimentos técnico-políticos, de modo que a realização dessas, sobretudo se significativamente abrangentes, não depende apenas do técnico ou gestor do programa, mas do contexto político e histórico em que a Política Pública se insere.

Se o programa e projeto produzem resultados e impactos, é necessário analisar os custos envolvidos na operacionalização de suas atividades, equipamentos e pessoal – *avaliação custo-efetividade*. O custo-efetividade das intervenções, isto é, o valor gasto para produzir unidades de resultados e impactos em um período de tempo e território específicos, é certamente uma informação fundamental para avaliar a manutenção dos programas e projetos no futuro e em outros contextos. Ademais, se bem-realizadas – ou seja, com contabilidade precisa de custos e vetor abrangente de indicadores de resultados –, tais avaliações fornecem parâmetros cruciais à comparação de diferentes intervenções sociais e informa gestores nas decisões técnicas e políticas acerca de oportunidades de melhoria de eficiência, continuidade, descontinuidade, ou mesmo a expansão de programas e projetos.

Outra tipologia relevante de Avaliação, é aquela que a diferencia segundo a natureza do produto informacional elaborado (JANNUZZI 2016a). Em uma perspectiva metodológica, os esforços de avaliação podem se estruturar em quatro tipos de produtos mais gerais, com maior ou menor aderência e especificidade ao problema social ou intervenção programática desenhada, a saber:

- *Estudos avaliativos* – análises com base em dados secundários ou compilação de artigos e trabalhos já realizados na temática, com maior ou menor abrangência;
- *Pesquisas de avaliação* – levantamentos primários, quali ou quantitativos, desenhados com objetivos de produção de evidências mais específicas e necessárias ao aprimoramento da intervenção;
- *Meta-avaliações* – recensões sobre estudos avaliativos, pesquisas e experiências nacionais, subnacionais e internacionais de programas e projetos implementados;
- *Relatórios-síntese de avaliação e portais web* – com informação mais sumária, na forma de indicadores gerais, para uma comunicação mais objetiva acerca de aspectos do diagnóstico, implementação e resultados dos programas e projetos.

Cada uma dessas modalidades tem sua utilidade frente ao público para o qual se quer atingir e, naturalmente, implica em tempo e recursos diferentes para sua produção. Dessas modalidades, vale destacar a Meta-avaliação, estratégia avaliativa que demanda recursos e prazos relativamente modestos. Ademais, com o aumento da produção de estudos e avaliações sobre políticas e programas públicos nos governos, centros de pesquisa, organismos internacionais e universidades tem sido possível produzir sínteses avaliativas e meta-avaliações como recursos para a análise de questões de implementação e para o balanço de resultados e de impactos de programas. Nas áreas da saúde e da educação, em especial, que têm políticas e programas mais antigos nos países da região e sobre as quais pesquisadores possuem maior conhecimento, tem havido uma produção volumosa de pesquisas nas universidades, em centros de pesquisa e nos ministérios.

Meta-avaliação, síntese avaliativa, revisão sistemática e meta-análise de estudos avaliativos são diferentes termos para denominar uma recensão estruturada sobre uma série de estudos e pesquisas de avaliação acerca de um programa ou um projeto específico, realizada por diferentes pesquisadores ou instituições, em momentos distintos, com propósitos, metodologias e amostras particulares. Com suas nuances e formas de apresentação específicas, essas ferramentas reúnem evidências sobre vários aspectos analisados do programa ou do projeto, em particular, sobre os principais resultados dos mesmos, as condições de sucesso, as boas e más práticas e os desafios de implementação. Trazem subsídios para definição de temas de investigação primária posterior, quali ou quantitativa, ao apontarem dimensões impactadas, perfil de beneficiários, contextos de implementação etc. Sistematizam diferentes perspectivas de julgamento e análise do programa, ao trazer à tona resultados de distintos avaliadores e instituições. Ajudam, desse modo, a organizar informações relevantes para avaliação de mérito de programas, ao permitirem encontrar pontos de convergência ou divergência em estudos produzidos anteriormente e oferecerem, ao decisor político, um contexto mais abrangente para tomada de decisão. Contribuem também para a formulação de roteiro de investigação em estudos qualitativos e questionários de pesquisa de campo, assim como para a identificação de dimensões tangíveis e latentes na análise de mérito de programas.

Na literatura da área, há autores que usam os quatro termos de forma intercambiável, mas há quem os diferencie, para designar trabalhos avaliativos de natureza diferente. Para alguns autores, o termo meta-avaliação significa uma avaliação da qualidade e robustez de um estudo avaliativo, tendo como referência um conjunto de critérios ou protocolos (ELLIOT, 2011). Leeuw (2011) comenta que os termos – e procedimentos – meta-análise e meta-avaliação surgiram mais ou menos ao mesmo tempo em meados dos anos 1970, para designar estudos de sistematização de achados em outras pesquisas – no primeiro caso – e análise da robustez metodológica das avaliações – no segundo. Contudo, ainda hoje, em parte da literatura internacional, meta-avaliação também pode ser usado como sinônimo de meta-análise, decisão assumida neste texto (DANIDA, 2004)⁴.

Uma das dificuldades na elaboração de uma meta-avaliação é o seu compromisso em explicitar, de forma organizada, os diversos achados substantivos e os pressupostos metodológicos de cada estudo analisado. Nem sempre é possível fazer isso, pelas diferenças de preocupações de pesquisa dos autores e, também, pela falta de detalhes metodológicos nos trabalhos. Quando possível, no entanto, pode se fazer um esforço de criação de categorias de sistematização e de quadros-síntese, para que fiquem claros os aportes de cada estudo avaliativo, as convergências em um determinado aspecto e os achados não convergentes. Sem essa preocupação com a apresentação estruturada dos estudos, a meta-avaliação pode não ser útil como recurso de reflexão avaliativa, mas introduzir mais ambiguidade e incerteza na avaliação.

O quadro-síntese de uma meta-avaliação deve conter os elementos fáticos que permitam, ao avaliador ou decisor, rapidamente entender os motivos de resultados aparentemente contraditórios entre estudos sobre um mesmo programa. Afinal, os achados diferentes decorrem de estudos realizados em momentos, regiões ou públicos distintos? Seriam produto de aplicação de técnicas, modelagem, pressupostos distintos na construção dos indicadores e na análise? Seriam resultado de diferentes linhas interpretativas ou valores públicos empregados nas avaliações?

Enfim, a meta-avaliação permite fazer uma avaliação “triangulada” de políticas e programas, contrapondo achados de pesquisas realizados com diferentes sujeitos, regiões, períodos, técnicas, fonte de dados, critérios avaliativos, paradigmas e perspectivas de análise. Em uma

perspectiva pós-positivista, a ciência social – e a avaliação – não está em busca de uma “verdade” incontestável, mas sim na sistematização de conhecimento técnico e científico coerente na lógica argumentativa, disciplinado na produção de evidências e robusto na perspectiva paradigmática adotada. Tomadores de decisão precisam conhecer, em bases plurais, os achados sobre a questão em tela e possíveis implicações de suas escolhas.

Etapas para especificação e realização de uma pesquisa de avaliação

A especificação de uma Avaliação envolve um conjunto de atividades organizadas em etapas sucessivas. Para fins didáticos, a proposta apresentada em Jannuzzi (2016a) parece suficientemente robusta e justificável para ilustrar os esforços diferenciados necessários no ciclo de produção da maior parte dos estudos avaliativos de programas públicos, de menor ou maior complexidade. Nessa proposta, a elaboração de pesquisas de avaliação de programas pode ser dividida em seis grandes etapas de atividades,. São elas:

- Estudo de Avaliabilidade do programa;
- Concepção e delimitação do escopo da avaliação;
- Desenho e especificação metodológica da pesquisa;
- Condução do trabalho de pesquisa documental e/ou de campo;
- Análise de dados da pesquisa e discussão de achados com a equipe do programa ;
- Produção de relatório e Disseminação de resultados.

Naturalmente o processo de elaboração não é tão linear e unidirecional como pode parecer à primeira vista essa sequência de etapas. Como em todo processo de produção de conhecimento, decisões tomadas em uma etapa anterior podem ter que ser revistas, de modo a tornar mais viável ou robusto a realização das etapas posteriores. Vale citar dois exemplos, que ocorrem com alguma frequência: a pesquisa de avaliação é definida com escopo abrangente, relevando-se inviável – em custos e tempo – na etapa do seu desenho metodológico; ou, definida a estratégia metodológica da pesquisa, ele se revela, em campo, difícil de ser operacionalizada pelas dificuldades logísticas encontradas. Nesses casos é necessário retomar a etapa anterior do planejamento da pesquisa.

⁴ Segundo Danida (2004), “a meta-evaluation is an instrument used to aggregate findings from a series of evaluations. It also involves an evaluation of the quality of this series of evaluations and its adherence to established good practice in evaluation” (p.11).

O Estudo de Avaliabilidade é uma etapa crucial do processo, já que define se a avaliação vai ser ou não realizada e que contornos deve assumir nas fases seguintes. Trata-se de analisar se a avaliação deve ser realizada, se ela pode produzir evidências relevantes para o programa em questão, se o estágio em que se encontra o programa justifica um esforço maior de investigação sobre seus componentes. Estaria o programa em estágio ou momento propício para uma investigação mais detalhada? Em que aspecto? No seu desenho de intervenção, em atividades de sua implementação ou já estaria avançado suficientemente para uma avaliação de resultados e impactos? O programa, de fato, já existe? Houve dotação orçamentária, repasses para agentes operadores e alocação de equipes? Já está implantado no país ou região em que fora previsto? Existe um sistema de gestão, registro de atividades e indicadores de monitoramento? Existem bases de dados para avaliação do programa? Já foram feitas avaliações anteriores do programa? Essas são perguntas básicas que permitem assegurar a factibilidade (ou não) de se fazer um esforço de avaliação de programa com possibilidades de produzir informações relevantes para seu aprimoramento, e qual deve ser o foco da pesquisa, considerando o estágio em que se encontra o programa no ciclo.

Antecipar a avaliação de um programa que se encontra em seus primeiros estágios de implantação pode levar a produção de resultados pouco relevantes ou consistentes. Pode, inclusive, desviar o foco e atrapalhar o processo de implantação das rotinas operacionais do programa e de resolução dos problemas básicos de qualquer novo projeto no setor público, como a obtenção dos recursos financeiros, contratação de equipes, articulação com outros setores do governo, criação de um sistema de gestão, de painel de indicadores de monitoramento, entre outros. Realizar uma avaliação de resultados e impactos sem se assegurar que as rotinas básicas do programa já se encontram estruturadas e operando adequadamente pode produzir evidências de baixa ou falta de efetividade do programa. De forma apressada pode-se deslegitimar uma boa e meritória proposta de intervenção que, por complexidade da problemática visada ou do arranjo operacional idealizado para mitigá-la, não teve seu tempo adequado de maturação. Por outro lado, postergar avaliações formativas de processos críticos pode conspirar para atraso da implantação do programa, adiando reformulação necessárias ao seu desenho original.

O Estudo de Avaliabilidade recupera os principais documentos descritivos e normativos do programa, sistematizando seus objetivos, públicos-alvo, lógica de intervenção, mapeamento de principais atividades e seus produtos, principais agentes públicos e privados envolvidos, tipos de serviços e entregas. Também procura identificar os sistemas de gestão, bases de dados e pesquisas de avaliação já realizadas anteriormente. Certamente requer entrevistas de técnicos e gestores, em geral, envolvidos na coordenação do programa. Idealmente, a finalização dessa etapa não deveria resultar apenas na produção de uma nota sintética com as evidências e conclusões acerca da viabilidade, pertinência e continuidade (ou não) da avaliação, mas uma documentação síntese atualizada do programa, quem sabe, com um Mapa de Processos e Resultados ou Modelo Lógico, recurso dialógico interessante para iniciar a discussão da próxima etapa. Poderia ensejar também a elaboração de uma Meta-avaliação de um programa parecido, no sentido discutido anteriormente, como forma de, inclusive, auxiliar nas próximas etapas.

Se o programa se revela em condições de ser avaliado, iniciam-se as atividades de Concepção, delimitação dos objetivos e o foco da pesquisa de avaliação. Como já mencionado nesse texto, programas públicos são empreendimentos complexos, com muitas atividades, operando em contextos de vulnerabilidade social e de capacidade de gestão muito heterogêneos. É preciso ter clareza do foco avaliativo, procurando definir que componentes e que aspectos do programa serão avaliados. A etapa em que o programa se encontra já aponta as demandas centrais de avaliação, como ilustrado na apresentação da tipologia de Rossi *et al* (2004) na seção anterior. Se o programa foi recém-implantado, há sempre necessidade de avaliar processos e atividades, e identificar gargalos, qualidade dos serviços prestados junto aos públicos-alvo. Mas se já está operando há muito tempo pode ser o caso de uma avaliação somativa, de análise das contribuições do programa na mitigação do problema que lhe originou. Mas mesmo avaliações formativas ou somativas podem abordar muitos aspectos da implementação ou efeitos (positivos ou negativos) do programa. Por que determinadas atividades não estão se desenvolvendo plenamente? Por que em algumas regiões o programa obtém mais resultados? Os serviços estão sendo prestados satisfatoriamente, de acordo com padrões de qualidade definidos? Qual a cobertura do programa? Quem está se beneficiando dele?

A equipe coordenadora do programa pode já ter algumas questões de interesse, às vezes muito específicas, de natureza formativa. O decisor estratégico, a quem cabe negociar recursos para o programa, pode ter outras demandas, do tipo de avaliações *ex-post*, visando colher argumentos técnicos para mais recursos. Diferentes equipes e gestores podem ter visões distintas do que deve ser avaliado, e a especificação do objetivo da avaliação pode levar mais tempo do que se supõe. Reuniões técnicas, dinâmicas de grupo com circulação prévia de relatórios parciais ajudam formar consensos e clarificar as questões centrais a investigar.

Em dado momento, até por necessidade de finalização do esforço de concepção, emerge o tema de maior interesse, relevância estratégica ou criticidade a ser analisado no programa. É o início do Desenho metodológico da avaliação. Nessa etapa, a terceira do processo de planejamento da avaliação, são formuladas as perguntas gerais a serem respondidas, os contextos territoriais ou arquetípicos de interesse e os sujeitos a serem abordados – gestores estaduais, técnicos municipais, beneficiários do programa ou usuário dos serviços, a população em geral. É necessário escolher a técnica de coleta mais adequada para responder às questões levantadas, a natureza e tamanho da amostra de sujeitos ou instituições a ser entrevistada, as questões éticas suscitadas, considerando os prazos e recursos disponíveis.

Quanto às técnicas, a Pesquisa Social oferece várias alternativas: a compilação de estudos anteriormente já realizados sobre questões similares, a busca de dados secundários, a observação e suas variantes – visitação, observação participante etc. –, as entrevistas – semiestruturadas ou estruturadas –, os grupos de discussão, as pesquisas de campo com questionários – com amostras probabilísticas ou intencionais –, e os desenhos quasi-experimentais. Estes métodos podem ser classificados em dois grandes campos, de acordo com o tipo de perguntas e estruturação dos instrumentos de coleta e análise das informações para respondê-las: métodos qualitativos e métodos quantitativos (MINAYO *et al.* 2005)⁵.

As pesquisas de campo podem ser mais estruturadas, como as *enquetes quantitativas com marco amostral probabilístico* – necessárias à produção de indicadores de dimensionamento de públicos-alvo ou inferência representativa quanto aos resultados dos programas –, ou com *amostras intencionais* – mais rápidas e menos custosas, mas com limitado poder de generalização dos seus resultados. Podem ser menos

estruturadas, mais exploratórias, *de cunho qualitativo*, como grupos de discussão ou entrevistas em profundidade, dirigidas a usuários e beneficiários de programas e projetos, assim como aos gestores e técnicos encarregados da operacionalização dos serviços. São especialmente importantes para identificar questões latentes acerca da implementação dos programas não antecipados, quando de seu desenho. Como observado anteriormente, em algumas situações, particularmente em estudos em que se interessa dimensionar efeitos ou impactos específicos de uma intervenção, pode-se aplicar um tipo particular de instrumento estruturado – a pesquisa com delineamento *quasi-experimental*.

Não existe um método ou estratégia “padrão-ouro” para a produção de uma avaliação. O melhor método é o que produz as evidências que respondem de forma consistente às demandas requeridas, ao tempo de seu uso na decisão da gestão pública. Informação precisa, mas produzida a custos e tempo não condizentes com a tempestividade da gestão; ou, ainda, informação rapidamente produzida, mas não consistente e robusta em termos metodológicos, certamente não se presta a orientar decisões cruciais acerca dos rumos de um programa ou projeto social. Perspectiva multidisciplinar de investigação, triangulação de métodos e de sujeitos entrevistados, esforços combinados de avaliação interna – com gestores e técnicos que conhecem os problemas e as atividades do programa e projeto – e de avaliação externa – com pesquisadores especializados e apoio de equipe de campo – é que garante a credibilidade e robustez necessárias ao aprimoramento da gestão e desenho das intervenções programáticas.

Da finalização do processo de especificação da metodologia da pesquisa de avaliação ao início do trabalho de campo da pesquisa é preciso definir como a pesquisa será efetivamente realizada: pela equipe interna do programa, com a contratação de equipe externa – instituto de pesquisa, firma de consultoria, consultores especializados ou fundação universitária-, ou por meio de um arranjo misto dos listados anteriormente. Contratações externas de pesquisas de avaliação, no setor público, envolvem esforços expressivos de definição dos termos de referência para contratação, para garantir que, ao final do processo, o relatório traga, de fato, os insumos esperados, relevantes e consistentes para sua apropriação junto às equipes.

⁵ Vide nesse sentido os manuais de técnicas de pesquisa social, sistematizados por GHEZZI, SANTOS (2017a, 2017b).

O Trabalho de Campo é quarta etapa do processo de planejamento da avaliação, outra fase crítica da mesma (é qual etapa não seria, afinal?!). Pela natureza operacional das atividades envolvidas o trabalho de campo pode não gozar do status que se atribui às outras etapas “mais nobres” da avaliação, desenvolvidas em escritório, mas não diminui de modo algum sua importância e a necessidade de rigor técnico e zelo na sua condução. Todo esforço de concepção e especificação metodológica pode ser colocado a perder se não forem observados os cuidados necessários com o treinamento de equipe, pré-teste de instrumentos, planejamento logístico da coleta, supervisão de campo e análise crítica das evidências coletadas na pesquisa. Fazer estimativas conservadoras de tempo e recursos para essas atividades pode ser um risco.

Se a avaliação não envolve extenso trabalho de campo, mas se constitui em um estudo avaliativo baseado na revisão de outros textos já produzidos, análise de dados secundários ou ainda uma meta-avaliação, é recomendável também ter um acompanhamento muito próximo e periódico, para que o técnico ou pesquisador envolvido não enverede por questões secundárias, às vezes de seu maior interesse, esquecendo-se de atender as demandas avaliativas que orientaram a especificação do estudo.

Pesquisas qualitativas têm, em geral, amostras com tamanhos menores, mas requerem entrevistadores de campo mais experientes, com conhecimento sobre a temática investigada, com capacidade de improvisação criativa e adaptação. Entrevistas em profundidade ou grupos de discussão podem desconcertar um entrevistador ou moderador pouco experiente. Levantamentos quantitativos envolvem esforços de coleta em amostras mais amplas, exigindo entrevistadores com prática de abordagem pessoal, capacidade de localização e disponibilidade para deslocamento pelo território. Em um caso e noutro, a equipe de campo contratada precisa estar comprometida com os objetivos da pesquisa, com a qualidade do trabalho e bem treinada nos conceitos e quesitos do instrumento de coleta.

O trabalho de campo não precisa ter terminado para se iniciar a Análise dos dados e Interpretação dos resultados da pesquisa, penúltima fase do planejamento da avaliação de programas. A exploração inicial dos dados coletados com a pesquisa ainda em campo pode, inclusive, ajudar a identificar eventuais problemas de coleta e corrigi-los, como comentado há pouco. Ademais, tal exploração

permite ir apreciando alguns achados e testando algumas das hipóteses aventadas quando da concepção da pesquisa de avaliação, adiantando tarefas de análise cujo tempo pode acabar sendo encurtado pelos atrasos da coleta de campo e a pressão das datas críticas de apresentação de resultados para equipe de coordenação do programa.

Tal como para coleta de informações em campo, há diversos métodos de pesquisa para auxiliar na análise dos resultados da pesquisa, ainda que as técnicas empregadas na coleta determinem, em boa medida, as abordagens analíticas que podem ser adotadas. Certamente não é a sofisticação da técnica analítica que confere legitimidade científica à análise ou relevância substantiva à pesquisa de avaliação. Não é o emprego de um modelo estatístico, econométrico ou análise quantitativa de dados qualitativos que pode garantir aportes relevantes e robustos da pesquisa. Há certa mitificação – senão, mistificação – das abordagens quantitativas complexas na pesquisa acadêmica em disciplinas das Ciências Sociais Aplicadas, que acabam influenciando as pesquisas de avaliação. É como se a complexidade do modelo quantitativo pudesse substituir a “imaginação sociológica” do pesquisador na busca das respostas para as perguntas de interesse, ou como se o emprego de tais técnicas pudesse conferir neutralidade na interpretação dos dados de pesquisas de avaliação.

Com avanço da análise dos dados e maturação interpretativa dos resultados inicia-se a sexta e última etapa do processo de planejamento da avaliação: a Produção do relatório final e definição das estratégias de disseminação dos resultados da pesquisa de avaliação. Idealmente, os resultados da pesquisa devem ser compartilhados, se consistentes, para as equipes da pesquisa e do programa, envolvidas na avaliação, desde o início do trabalho de campo e prossegue durante o processo. É uma boa estratégia para, gradualmente, preparar a equipe do programa acerca dos resultados mais incisivos e negativos ao final, e receber contribuições para interpretação dos resultados.

Como advogam Imas e Rist (2009), o relatório de pesquisa deve ser claro, organizado por temas gerais ou de acordo com as grandes perguntas de interesse da avaliação. Sua finalidade é facilitar que os resultados e os destaques sejam entendidos para a equipe do programa, seus operadores e seus formuladores. O objetivo do relatório deve ser o de comunicar os achados aos seus principais interessados, não impressionar a comunidade de avaliadores pela complexidade das técnicas ou ineditismo da abordagem metodológica adotada, critérios

que presidem, muitas vezes, a elaboração de relatórios de pesquisa. Não deve se pretender exaustivo na apresentação dos resultados em um único documento. Organizar os resultados segundo relatórios temáticos pode ser uma boa estratégia, sobretudo quanto se trata de avaliações somativas, que abordam muitas problemáticas diferentes, em grande profundidade analítica.

O entendimento de pesquisas de avaliação como instrumentos de aprendizagem organizacional e de aporte de conhecimentos para inovação de programas requer esforços intensos e criativos para disseminação de seus resultados, para todos os seus potenciais usuários, dos “decisores” estratégicos aos operadores dos programas na ponta, em contato com os usuários e beneficiários. É necessário também ter produtos de divulgação para sociedade, fiadora e financiadora em última instância dos programas de das pesquisas de avaliação. Os produtos de disseminação de resultados precisam estar, pois, ajustados à linguagem dos vários públicos visados, envolvendo diferentes meios e estratégias. Notas em boletins de circulação entre operadores de programas, participação em eventos com técnicos de programas, em congressos científicos são algumas dessas iniciativas a serem planejadas.

Considerações finais

O paradigma da “Gestão-por-Resultados” e “Política-Pública-orientado-por evidências” teve grande influência na formação técnica da comunidade epistêmica e na institucionalização de práticas da Avaliação no Brasil e América Latina, sobretudo nas escolas de economia e áreas de finanças e controle da Administração Pública. Certamente com contribuições meritórias, essa perspectiva hegemônica tem, por outro lado, se prestado a um processo de forte “desacreditação” de políticas e programas públicos. Mais do que apontar problemas, gargalos, bons e maus resultados para públicos e contextos específicos, avaliações lastreadas sob princípios de eficiência e conformidade administrativa têm sido usadas para legitimar o corte de recursos e descontinuidade de políticas e programas com forte legitimidade institucional e social.

Mais recentemente tem se observado um esforço de produzir Avaliações segundo outras perspectivas teórico-metodológicas e

políticas na região. À semelhança do processo histórico de construção do Estado de Bem-Estar Social nos países centrais ao longo do século XX, tem se procurado propor conceitos, modelos e práticas que possam, de fato, subsidiar o processo de desenho, implementação e análise de resultados das emergentes Políticas e Programas públicos no Brasil e outros países. Na encruzilhada histórica em que se encontra a América Latina, a avaliação de políticas e programas precisa se orientar pelos valores públicos que foram restituídos no contexto de democratização e institucionalização de direitos nos últimos 30 anos. Sem tal compromisso, as avaliações não aportarão contribuições relevantes para construção de uma sociedade mais próspera e justa.

Bibliografia

DANIDA. **Meta-Evaluation Private and Business Sector Development Interventions**. Denmark, 2004.

ELLIOT, L.G. Meta-avaliação: das abordagens às possibilidades de aplicação. **Ensaio: aval. pol. públ. Educ.**, Rio de Janeiro, v. 19, n. 73, p. 941-964, out./dez. 2011.

GERTLER, P. et al. **Avaliação de impacto na prática**. Washington, Banco Mundial, 2015.

GHEZZI, D.R.; SANTOS JR., J. **Métodos de pesquisa em Ciências Sociais: Bloco Qualitativo**. Cebrap/Sesc, São Paulo, 2017a.

GHEZZI, D.R.; SANTOS JR., J. **Métodos de pesquisa em Ciências Sociais: Bloco Quantitativo**. Cebrap/Sesc, São Paulo, 2017b.

GUSSI, A. Outras epistemologias e metodologias: a experiência do Mestrado de Avaliação de Políticas Públicas. **Revista Aval**, Fortaleza, 2(16):168-183, 2019.

GUSSI, A.F.; OLIVEIRA, B. R. Discutindo paradigmas contra-hegemônicos de avaliação de políticas públicas. In. **Anais do Encontro Nacional de Ensino e Pesquisa do Campo de Públicas**, Brasília, 2017.

IMAS, L.G.M.; RIST, R. **The road to results: designing and conducting effective development evaluations.** Washington: World Bank, 2009.

JANNUZZI, P.M. **Monitoramento e avaliação de programas sociais: uma introdução aos conceitos e técnicas.** Campinas: Alínea, 2016a.

JANNUZZI, P.M. Eficiência econômica, eficácia procedural ou efetividade social: Três valores em disputa na Avaliação de Políticas e Programas Sociais. **Desenvolvimento em Debate**, v.4, n.1, 2016b, p. 117-142.

JANNUZZI, P.M. Mitos do Desenho Quase-experimental na Avaliação de Programas. **Revista NAU Social** – v.09, n.16, p. 76 – 90, 2018.

JANNUZZI, P.M. **Informação, estatística e políticas públicas no Brasil:** contribuições de pesquisas do IBGE para as políticas de desenvolvimento social e combate à fome (2004-2014). Monografia (Especialização em Política e Sociedade) – Instituto de Estudos Sociais e Políticos, Universidade do Estado do Rio de Janeiro, 2020.

LEEUW, F.L. On the contemporary history of experimental evaluations and its relevance for policy making. **Comparative Policy Evaluation**, 15:11-26, 2011.

MCDONALD, J. HAWTHORN, L. **Program Evaluation and Performance Measurement: An Introduction to Practice.** Thousand Oaks; Sage, 2006.

MINAYIO, C. *et al.* **Avaliação por triangulação de métodos: abordagem de programas sociais.** Rio de Janeiro: Ed. Fiocruz, 2005.

PAES-SOUZA, R.; VAITSMANN, J. Avaliação de programas e profissionalização da gestão pública. **Revista Brasileira de Monitoramento e Avaliação**, Brasília, n. 1, p. 18-35, 2011.

PATTON, M.Q. **Utilization-focused evaluation: the next century text.** Thousand Oaks, Sage, 1997.

RAVALLION, M. Should the randomists rule? **Bekerley Electronic Press**, p.1-5, fev. 2009.

ROSSI, P. *et al.* **Evaluation: a systematic approach.** Thousand Oaks: Sage, 2004.

SAGI. **SAGI: informação e conhecimento para as políticas de desenvolvimento social.** 1ed. Brasília, 2016.

WORTHERN, B.R. *et al.* **Avaliação de programas: concepções e práticas.** São Paulo, Ed.Usp/Gente, 2004.