

✓ A implementação no centro da Avaliação de Políticas Públicas¹

Paulo de Martino Jannuzzi²

¹ Texto revisado a partir de versões elaboradas anteriormente para exposição em Mesa Redonda no Seminário Internacional de Avaliação da Escola Nacional de Administração Pública, Brasília, em 20-21 de março de 2018 e no 14º Seminário Internacional de Avaliação, do GIF/FRM/ItauSocial, São Paulo em 15/08/2018. Para uso no Curso de Avaliação Sistemática na Fundação Carlos Chagas, outubro de 2018.

² Professor da Escola Nacional de Ciências Estatísticas do Instituto Brasileiro de Geografia e Estatística (ENCE/IBGE), professor da Faculdade Cesgranrio-RJ e Pesquisador CNPq.

Resumo: Com a ampliação de escopo e escala das Políticas Públicas desde a Constituição Federal de 1988, o campo de avaliação vem se consolidando em diferentes contextos institucionais no país. Em que pesem os avanços na produção de conhecimento na área, há certo desencanto acerca do papel e uso da avaliação. Este texto procura trazer subsídios para mitigar a frustração de gestores e avaliadores com a avaliação de programas e/ou sua apropriação no processo decisório, especialmente no contexto desafiador de estruturação de políticas públicas no Brasil. O texto inicia-se com uma discussão sobre a complexidade e estágio de maturidade das Políticas Públicas no Brasil. Passa-se então à exposição da necessidade de uma perspectiva sistêmica de avaliação e, em seguida, de colocar a Implementação no centro das preocupações com a Avaliação. Finaliza-se advogando-se pelo uso de critérios abrangentes e republicanos na apreciação de mérito das políticas e programas no país.

Palavras-chave: Avaliação, abordagem sistêmica, Políticas Públicas, Implementação

EN

Abstract: With the expansion of the scope and scale of Public Policies since the Federal Constitution of 1988, the field of evaluation has been consolidating in different institutional contexts in the country. Regardless of the knowledge advances in the area, there is some disenchantment about the role and use of evaluation. This text seeks to provide subsidies to mitigate this frustration of managers and evaluators with the evaluation of programs and / or their appropriation in the decision-making process, especially in the challenging context of public policies making in Brazil. The text begins with a discussion about the complexity and stage of maturity of Public Policies in Brazil and, after, the need for a systemic evaluation perspective. It is then argued that Implementation should be at the center of Evaluation concerns. It ends by advocating for the use of comprehensive and republican criteria in assessing the merits of policies and programs in the country.

Key-words: Evaluation, systemic approach, Public Policies, Implementation

ES

Resumen: Con la ampliación de alcance y escala de las Políticas Públicas desde la Constitución Federal de 1988, el campo de evaluación se viene consolidando en diferentes contextos institucionales en el país. En que pesen los avances en la producción de conocimiento en el área, hay cierto desencanto acerca del papel y uso de la evaluación. Este texto busca traer subsidios para mitigar esa frustración de gestores y evaluadores con la evaluación de programas y / o su apropiación en el proceso decisório, especialmente en el contexto desafiante de estructuración de políticas públicas en Brasil. El texto se inicia con una discusión sobre la complejidad y la etapa de madurez de las Políticas Públicas en Brasil. Se pasa a la exposición de la necesidad de una perspectiva sistémica de evaluación y, a continuación, de colocar la Implementación en el centro de las preocupaciones con la Evaluación. Se finaliza abogándose por el uso de criterios amplios y republicanos en la apreciación de mérito de las políticas y programas en el país.

Palabras-clave: Evaluación, enfoque sistémico, Políticas Públicas, Implementación

FR

Résumé: Avec l'élargissement de la portée et de l'échelle des politiques publiques depuis la Constitution fédérale de 1988, le domaine de l'évaluation s'est consolidé dans différents contextes institutionnels du pays. Indépendamment des progrès de la production de connaissances dans la région, le rôle et l'utilisation de l'évaluation suscitent un certain désenchantement. Ce texte cherche à fournir des subventions pour atténuer cette frustration de gestionnaires et évaluateurs avec l'évaluation des programmes et / ou leur appropriation dans le processus de prise de décision, en particulier dans le contexte difficile de formulation des politiques publiques au Brésil. Le texte commence par une discussion sur la complexité et le stade de maturité des politiques publiques au Brésil. La nécessité d'une perspective d'évaluation systémique est alors exposée, puis mise en place au centre des préoccupations de l'évaluation. Elle se termine par un plaidoyer pour l'utilisation de critères globaux et républicains dans l'évaluation des avantages des politiques et des programmes dans le pays.

Mots-clés: évaluation, approche systémique, politiques publiques, mise en œuvre

1. Introdução

A Avaliação de programas sociais tem se fortalecido no Brasil nos últimos 30 anos, com a ampliação do escopo e cobertura das políticas públicas. A promulgação da Constituição Brasileira de 1988 instituiu um conjunto amplo de direitos sociais – como acesso universal e integral à segurança alimentar, saúde, assistência social, educação básica, intermediação e qualificação profissional – assim como estabeleceu princípios básicos de atuação do Estado no combate à pobreza, à desigualdade econômica e à discriminação de segmentos populacionais – levando a estruturação de programas redistributivos de transferência de renda e ações afirmativas para negros e mulheres.

Tem se vivenciado no país o que se passou em países desenvolvidos há mais de 50 anos quando, do pós-guerra até anos 1970, com a estruturação de Políticas Públicas nesses países – e do Estado de Bem-estar suas várias modalidades – o campo de Avaliação desenvolveu-se fortemente. No caso dos Estados Unidos, por exemplo, em que esse processo se deu mais tardiamente, Patton (1997) comenta que o fortalecimento do campo da Avaliação se deu nos anos 1960 com os massivos recursos federais alocados nos programas sociais, ligados ou decorrentes do Plano “Great Society”. O investimento federal em programas voltados à eliminação da fome e pobreza ampliou-se para outras áreas de políticas – habitação, emprego, renovação urbana – requerendo um esforço mais abrangente de avaliação da efetividade da ação governamental.

O campo profissional, até então fortemente centrado na avaliação educacional, passa a ter demandas mais complexas e multidisciplinares de avaliação (WORTHERN et al., 2004). E para além da apreciação de resultados e impactos, para uso de decisores de nível estratégico, as avaliações passam a despertar interesse crescente de gestores em níveis médios da Administração Pública, preocupados com as dificuldades de implementação dos programas públicos. Novos profissionais, especialistas e cientistas sociais passam a integrar a comunidade de avaliadores, trazendo novas abordagens metodológicas e técnicas para o campo (PATTON 1990, ROSSI et al 2004).

De fato, processo semelhante se tem observado no Brasil, em partir desde os anos 2000, como mostram a produção de estudos sobre políticas públicas nas universidades e centros de pesquisa (IPEA, SEADE etc) e como relevam as avaliações realizadas ou contratadas pelos Ministérios setoriais (JANNUZZI 2016a). O caso das avaliações realizadas pela Secretaria de Avaliação e Gestão da Informação (SAGI) do Ministério de Desenvolvimento Social e Combate à Fome

(MDS), sobretudo no âmbito dos esforços de mitigação da pobreza e vulnerabilidade social articulados no Plano Brasil Sem Miséria (BSM)¹, é paradigmático nesse sentido, como descrito em documento de balanço de uma década desse esforço avaliativo (SAGI 2016). A ampliação da investigação social nas pesquisas do IBGE, dos instrumentos de avaliação educacional do INEP e dos sistemas de informação do Ministério da Saúde são outros casos ilustrativos de como a informação e avaliação passaram a cumprir papel cada vez mais relevante na gestão das políticas públicas. Diagnósticos sobre públicos-alvo dos programas, indicadores de monitoramento, avaliações de processos de implementação das políticas e análises mais abrangentes de resultados e impactos passaram a ser encontrados com maior regularidade no governo federal e também em vários estados e municípios de maior porte.

Em que pese esses avanços no campo da Avaliação no país, ainda persistem dificuldades e “desencantos”, seja por parte de demandantes e usuários de avaliações, seja de avaliadores envolvidos no processo. Como já registrado por Cotta (1998) há algum tempo, há quem diga que, na prática, as avaliações não subsidiam o processo decisório porque seus resultados são inconclusivos, inoportunos e irrelevantes. São inconclusivos em função das limitações de recorte analítico ou metodológico que os avaliadores adotam na análise de realidades complexas; inoportunos devido ao “tempo acadêmico” que levam para produzir evidências frente ao “tempo da tomada de decisão da gestão”; irrelevantes porque não respondem às perguntas que, de fato, são importantes para apreciação do mérito ou para aprimoramento da intervenção. Por outro lado, há que diga que os gestores e usuários não entendem a profundidade analítica e alcance dos estudos avaliativos, que eles estão mais interessados nos “bons resultados” e não nos “achados obtidos mediante às mais rigorosas técnicas” e enfim, que eles não se apropriam das avaliações no processo decisório.

Este texto procura trazer subsídios para mitigar essa dupla frustração – de gestores e avaliadores – com a avaliação de programas e/ou sua apropriação no processo decisório, especialmente no contexto desafiador de estruturação de políticas públicas no Brasil. Nesse sentido, o texto está estruturado em três seções, além dessa introdução e das conclusões. Inicia-se com uma discussão sobre a complexidade das políticas públicas, passa-se à perspectiva sistêmica de avaliação e, por fim, a necessidade de colocar a Implementação no centro das preocupações com a Avaliação no país, contribuição mais central desse trabalho. À guisa de conclusão advoga-se brevemente pela necessidade de utilização de critérios e valores mais abrangente para apreciação de mérito de políticas e programas.

¹ O Plano BSM foi uma estratégia intersectorial e interfederativa coordenada pelo MDS entre 2011 a 2015, articulando mais de uma centena de ações e programas federais, com objetivo de eliminação da fome e da extrema pobreza, ampliação de acesso a serviços públicos e a programas de inclusão produtiva, destinado às famílias de mais renda e vulneráveis no país, inscritos no Cadastro Único de Programas Sociais. Para mais detalhes vide livro de balanço do plano em Campello, Falcão e Costa (2014).

2. Avaliação e Tomada de Decisão em contexto de complexidade de políticas

Desencanto com o uso efetivo das avaliações, como o descrito na Introdução, tem sido registrados na bibliografia internacional do campo, como revelam Patton (1997), MacDavid & Hawthorne (2006) e Imas & Rist (2009). Em caso extremo, o ceticismo acerca da utilidade dos estudos avaliativos- e dos próprios programas sociais- chegaram ao ponto de Rossi (1987) conferir-lhe status de “regularidade científica” na forma de “Leis Metálicas da Avaliação”. Das quatro “Leis”- de Ferro, Aço, Bronze e Zinco- do autor, vale parafrasear a última: somente programas sociais que tem chance de não dar certo são avaliados; se existem programas sociais com alguma efetividade, seus efeitos devem ser tão óbvios que as avaliações nem são necessárias².

Há certamente, de um lado e de outro, críticas consistentes, mas também expectativas muito exacerbadas sobre o “poder iluminista” das Avaliações, sobre sua “primazia” no processo decisório e ainda a falta de clareza sobre os valores que legitimam a proposição, desenho e avaliação das políticas públicas. Afinal, de um lado, a produção de conhecimento tem muitos limites “explicativos” em qualquer área das ciências, sejam elas ciências duras, macias ou moldáveis, como o campo de conhecimento técnico-científico-aplicado das Políticas Públicas. Não só não existe um só modo de fazer Ciência ou um único método “padrão-ouro” de produção e legitimação de conhecimento. Também não existem “verdades absolutas e incontestes” nas Ciências, quando mais nas Ciências Sociais e na Avaliação. Tais “Verdades” é o que se busca em algumas religiões; nas Ciências Sociais e Avaliação se busca achados e interpretações consistentes sobre os mesmos ou ainda, em uma perspectiva “latouriana”, fazer ciência é construir narrativas persuasivas e convincentes sobre os achados analisados, reconhecidas como legítimas e razoáveis pelas comunidades epistêmicas a que se pertence (KROPF, FERREIRA 1998).

De outro lado, avaliadores, por mais competentes e cuidadosos que sejam na produção de seus achados, não dispõem da legitimidade política e visão estratégica do decisor público. Recomendações de avaliações podem implicar necessidade de mais recursos orçamentários ou pessoal técnico, mudanças na legislação ou na pactuação com outros agentes ou ainda reformulações no desenho operacional da intervenção. As sugestões advindas de uma avaliação precisam ser analisadas em termos de sua viabilidade política e técnica, para não se correr o risco de interromper atividades que atendem a sociedade, ainda que não satisfatoriamente (CEJUDO 2010). Algumas comunidades de avaliadores imaginam, ingenuamente por vezes,

que seus diagnósticos e evidências, elaboradas mediante técnicas auto-proclamadas como “as mais robustas dentre as existentes”, apresentam as soluções “mais racionais” e definitivas para os problemas críticos que programas enfrentam.

Avaliações produzem indicadores, evidências qualitativas, resultados para informar o bom e plural debate, esclarecer aspectos positivos e negativos, considerando as limitações do escopo investigado e dos instrumentos de levantamento empregados. Ou seja, avaliador avalia; decisor decide. É sempre bom lembrar que nas democracias, ao contrário de uma tecnocracia ou plutocracia, a decisão em políticas públicas é eminentemente política, orientada em tese primariamente por princípios republicanos, constitucionais e valores humanos, e secundariamente- e desejavelmente-, por evidências técnicas sobre eficiência, eficácia e efetividade acerca das políticas e programas. Avaliações, como o desenho de políticas e programas, não são neutras, mas refletem perspectivas e valores de quem os realiza (WEISS 1993, PORTER 1995, YANES 2009). Resultados de um estudo avaliativo dependem das escolhas metodológicas realizadas e também das finalidades políticas mais gerais a que ele se presta. Concepções diversas da natureza da política pública moldam escolhas de critérios avaliativos, de técnicas de investigação e, naturalmente, o tipo de evidências e indicadores a produzir (JANNUZZI 2016a).

Superar esses “desencantos” requer o entendimento acerca da complexidade dos arranjos de implementação dos programas públicos e das necessidades de informação requeridas para tomada de decisão sobre os mesmos. É preciso reconhecer que os programas públicos mais impactantes foram implantados há menos de 20 anos, em contexto de grande heterogeneidade de capacidade de gestão e de vulnerabilidade de públicos-alvo pelo território. O rápido alargamento de escopo setorial e de cobertura populacional das políticas públicas, legitimado pela vontade majoritária da sociedade brasileira nas urnas no período, foi acompanhado de “deseconomias” inerentes à implementação nessa escala: problemas de articulação federativa e financiamento, fragmentação setorial e qualidade dos serviços na ponta. Nos países centrais, a estruturação do Estado de Bem Estar levou pelo menos o dobro do tempo, vale registrar.

O fato é que, no Brasil, políticas e programas abarcam um conjunto amplo de atividades e processos de trabalho, operando muitas vezes em vários ministérios e envolvendo milhares, senão milhões de agentes públicos e sociais, nas três esferas de governo. O país é bastante heterogêneo e desigual em termos socioeconômicos, com públicos-alvo muito diferentes de Norte a Sul do país, entre residentes nas capitais, periferias e interior. A capacidade de gestão e operação dos programas é igualmente diversa nesses contextos.

² Em reconhecimento ao autor, e para salvação e esperança dos avaliadores, ele acrescenta “if designing successful programs is a difficult task”, são as avaliações que podem ajudar a definir programas sociais melhores e mais efetivos (p.4).

Nos estados e municípios, em diversas políticas e programas, há problemas de alocação de pessoal com perfil adequado e em número suficiente, com formação em protocolos de atendimento de serviços, com disponibilidade de materiais e edificações adequadas, com dificuldades de relacionamento intergovernamental, para citar alguns dos principais problemas existentes.

3. Por uma perspectiva sistêmica de avaliação

Dado o quadro descrito anteriormente, para a maior parte das políticas no país, a informação que pode fazer diferença significativa para gestão não é, ao contrário do que advogam certas comunidades epistêmicas do campo, a produzida por rígidas avaliações de impacto³. Pelo estágio de maturação em que se encontram as políticas públicas no país, avaliações diagnósticas e avaliações de processo continuam sendo cruciais para desenho e aprimoramento de programas mais adequados às características dos públicos-alvo, às diferentes e capacidades de gestão pelo território e para identificação dos gargalos operacionais que mitigam a efetividade da ação governamental. Naturalmente, é preciso também produzir informação que evidencie o acesso e cobertura dos públicos-alvo almejados, que revele a qualidade dos serviços oferecidos aos mesmos, que aponte a observância aos princípios de equidade e coesão social inerentes à ação pública no país.

Avaliação de resultados deve ser agendada no momento devido, depois de resolvidos os diversos problemas operacionais que insistem aparecer na introdução de um novo programa, pelas falhas de planejamento, deficiências de alocação de recursos, adesão das equipes na ponta dos serviços, disputas políticas, para citar alguns dos problemas recorrentes em contextos complexos. Como bem coloca Patton (1997) :

A menos que se saiba que o programa está operando conforme o planejado, há poucos motivos para se esperar que ele esteja produzindo os resultados almejados. Mais ainda, até que o programa seja implementado e um “tratamento” esteja de fato em operação, há pouco sentido em avaliar impactos (p.200).

Tratamento deficiente não cura um paciente; se houve um diagnóstico correto, um tratamento médico efetivo certamente tem mais chances de curá-lo. Mas isso significa a execução zelosa de procedimentos, seguindo a receita definida e acompanhamento dos efeitos esperados e os colaterais no paciente, durante período

que pode variar pela gravidade da doença. Valendo-se da metáfora, um programa “em tratamento de fato” significa que os processos e atividades especificados no modelo lógico estão fazendo chegar os produtos, nas especificações requeridas, para os públicos-alvo do programa, em todos os seus contextos de operação. Ora, não há como garantir a conformidade na implementação do programa se avaliações que verifiquem tal adequação não forem realizadas. O fato de o recurso orçamentário ter saído de uma esfera administrativa para outra, que técnicos haviam sido contratados e material comprado não significava que o “Tratamento” havia iniciado ou já que começava a produzir os efeitos na intensidade que o “Clínico Geral” havia imaginado.

Não foram poucas as situações, segundo o autor, que avaliações de resultados e impactos de programas educacionais e sociais nos anos 1960-70 nos EUA não constataram qualquer impacto pelo fato de que eles padeciam de sérios problemas de implementação ou pelo fato de dela haver sequer começado. Propostas meritórias foram descartadas ou deixaram de receber recursos para expandirem sua cobertura por erro de especificação da avaliação – do tipo e momento- e não por erro do programa ou do “tratamento”. Tivessem os programas sido avaliados em perspectiva processual, em diferentes contextos de operação, mais simples, mais complexos, de maior e menor capacidade de gestão, talvez não só sobrevivessem a uma avaliação somativa no momento adequado, como poderiam ter uma expansão mais exitosa, ao apontar os ajustes necessários para funcionamento em cada âmbito.

Diferentemente da pesquisa acadêmica, a pesquisa avaliativa se legitima, segundo Demo (2001), por meio de dois princípios essenciais: a relevância política do estudo e a consistência metodológica do mesmo. A primeira refere-se ao grau de legitimidade, pertinência ou oportunidade da avaliação para subsidiar decisões políticas acerca dos programas; a segunda diz respeito à adequação dos instrumentos, técnicas e métodos empregados para responder às demandas de informação requeridas. Se a pesquisa acadêmica pode levantar questões de interesse particular do pesquisador e de sua comunidade epistêmica, a pesquisa avaliativa deveria ter compromisso precípuo de investigar questões efetivamente relevantes para aprimoramento das políticas.

Essa abordagem de avaliação, convergente com as preocupações de Patton (1997), Rossi et al (2004), Minayo et al (2005) e com as práticas de avaliação de programas sociais na SAGI/MDS, é o que se tem denominado de Avaliação Sistêmica (JANNUZZI 2016a). Políticas e programas são componentes de sistemas mais amplos e, em si, subsistemas, com seus subcomponentes específicos. Em uma perspectiva sistêmica de avaliação, para atender as diferentes

³ Banco Mundial e o Banco Interamericano têm, infelizmente, contribuído para incentivar essa perspectiva avaliativa, calcada certamente nos pressupostos de Estado Liberal que comungam. Programa bom nessa perspectiva é programa eficiente, não necessariamente efetivo em termos de cobertura de públicos, equidade de seu desenho e que revele outros impactos mais abrangentes como contribuição para sustentabilidade ambiental, respeito à diversidade identitária-cultural etc. Vide discussão nesse sentido em Jannuzzi (2016b).

necessidades de informação dos diferentes usuários e gestores envolvidos em políticas e programas públicos, deve-se realizar estudos diagnósticos *ex-ante*, avaliações de processo e análises *ex-post* de resultados e impactos – com indicadores quantitativos e também evidências qualitativas. Não há que se falar, pois, em avaliação, mas avaliações, ou melhor um conjunto de avaliações proposta na forma de plano, a ser executado em alguns meses e anos, conforme a complexidade da intervenção.

Nesse sentido, a Avaliação Sistêmica se contrapõe a outras perspectivas avaliativas, concebidas segundo modelos desenvolvidos para projetos isolados, de alcance limitado em termos de cobertura populacional e territorial, como se fossem experimentos controlados, e orientados a identificar se possíveis efeitos da intervenção podem ser, de fato, atribuíveis a ela. Naturalmente, os resultados e impactos considerados nessa perspectiva tendem a ser aqueles mais facilmente expressos em termos quantitativos. Mas, ao contrário de projetos sociais, políticas públicas são desenhadas para atender múltiplos objetivos, parte deles difíceis de se medir em uma escala razão. E quando várias políticas públicas estão operando conjuntamente torna-se ainda mais difícil atribuir a um programa específico, o eventual efeito produzido pelo conjunto (BAMBERGUER et al 2016). Ademais, essas avaliações estão centradas em medir impactos médios, mesmo que o desenho das intervenções tenham sido modificadas para atender públicos-alvo diferentes. Modificações nos desenhos de programa e características diferentes dos públicos-alvo podem implicar expectativas de resultados distintos em intensidade, natureza e tempo de maturação.

Nessas situações, uma estratégia avaliativa voltada a dimensionar contribuições e sinergias entre políticas e programas na produção dos resultados parece fazer mais sentido do que outras técnicas que, para estimar efeitos marginais atribuíveis a cada intervenção, encapsulam a complexidade do objeto avaliado por meio de amostras de Caso e Controle artificialmente construídas, valendo-se de procedimentos éticos por vezes questionáveis. O viés de seleção amostral que os “randomistas”⁴ imputam aos estudos avaliativos não experimentais é, pois, uma realidade também nas suas próprias pesquisas. Por que seriam os resultados de estudos experimentais mais “robustos” e “válidos” que outros de natureza metodológica diversa para inferência do mérito de um programa, quando as amostras em que se baseiam derivam de cadastros ou listas de seleção parciais e incompletos, que estão longe de representar, de fato, os potenciais públicos-alvo a que se destinam os serviços ou benefícios de um programa público?

Mais ênfase na contribuição integrada e menos na atribuição

marginal; amostras mais representativas da realidade complexa da intervenção que as intencionalmente escolhidas para conformidade da aplicação da técnica; mais equilíbrio entre a validade externa e validade interna parecem ser os princípios metodológicos a serem observados na condução de avaliações de políticas e programas. Problemas multicausais ou multideterminados requerem intervenções complexas que, por sua vez, requerem avaliações compreensivas. Demandam avaliações que sejam abrangentes em termos das informações relevantes para o estágio do programa no seu ciclo de implantação, que abordem diferentes sujeitos e agentes envolvidos na oferta e usufruto dos seus serviços e benefícios. Requerem, para isso, pesquisas e estudos que se valham de métodos mistos para produção das evidências empíricas e que, por fim, as articulem mediante uma reflexão crítica e triangulada para interpretação dos resultados.

Dados não falam por si, mas pelas diferentes linhas interpretativas adotadas. Expor diferentes interpretações dos resultados observados é antes uma postura científica e pessoal honesta que uma inconsistência teórica ou fragilidade metodológica de um estudo avaliativo. Se robustez metodológica é pressuposto da boa prática de produção de conhecimento técnico e científico, é preciso sensibilidade analítica para desvendar a complexidade inerente às problemáticas sociais e programas públicos. Adequação- e não ortodoxia- metodológica é a chave para produzir evidências relevantes na avaliação⁵.

4. A Implementação no centro da Avaliação das Políticas Públicas

Em uma perspectiva sistêmica, as dificuldades de implementação dos programas – que ocupam a maior parte do seu tempo de vida – deveriam estar no centro de preocupação dos gestores e avaliadores (Diagrama 1). É preciso redesenhar o esquema clássico do Ciclo de Políticas, de um circuito circular e sequencial de etapas de Demanda Social-Agenda Política- Formulação-Implementação-Avaliação, para um mais realístico para os gestores de programas públicos- e avaliadores- em que se reconhece a centralidade da Implementação no processo. É aí que os programas “vivem” a maior parte do tempo e onde correções, aperfeiçoamentos ou inovações incrementais determinarão o sucesso ou fracasso da intervenção pública (GUSSI; OLIVEIRA 2017). Afinal, a efetividade de um programa público depende, para além de um bom diagnóstico inicial- da demanda social, do contexto de operação e dos públicos-alvo- e de formulação de um

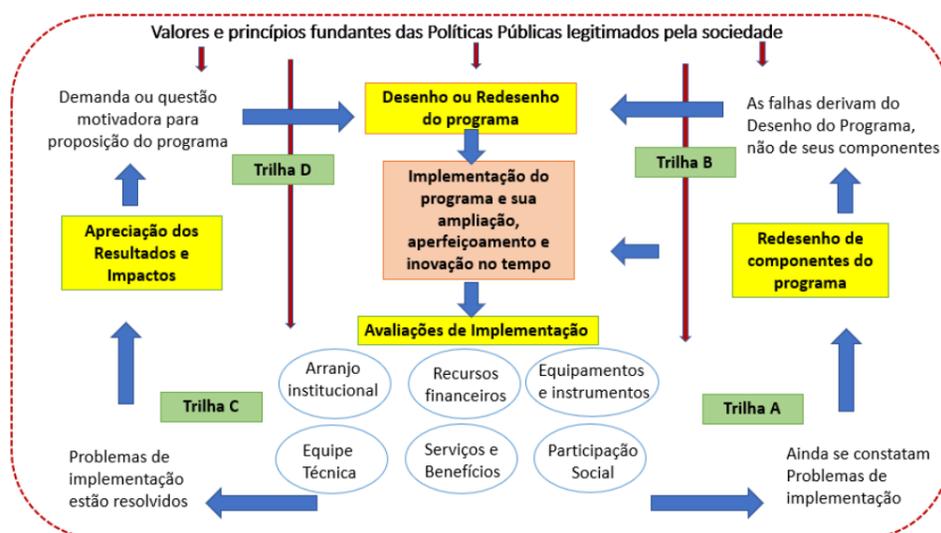
⁵ Michael Patton, em workshop posterior ao 14º Seminário Internacional de Avaliação, em 16/08/2018, trouxe algumas metáforas brilhantes para ilustrar a importância da sensibilidade do avaliador na escolha do método mais robusto e/ou adequado para contribuir para o usuário demandante. Uma delas é a do martelo e da chave de fenda: a robustez do primeiro presta-se a situações que requerem força bruta para atingir seus objetivos; o segundo, exige habilidade motora mais fina, em situações de maior fragilidade do objeto a ser tratado. Extrapolando a metáfora, um prego de aço entra em qualquer madeira, mais densa ou não, desde que batidas robustas sejam dadas; mas se você dispõe de um estojo com chaves de fenda de diferentes tamanhos, você pode escolher a mais ajustada à densidade da madeira e tamanho do parafuso a ser colocado. Outra metáfora é do GPS e a triangulação: para obter uma localização de latitude e longitude precisa através de um aparelho de GPS não basta um ou dois satélites, mas são necessários 3 satélites; da mesma forma, resultados mais precisos e consistentes de avaliação requerem a triangulação de pesquisas, sujeitos entrevistados e de achados.

⁴ Este é o termo com que Ravaillon (2009) denomina essa comunidade epistêmica de avaliadores que acreditam que somente avaliações experimentais são “robustas” e cientificamente legítimas na avaliação de programas e projetos sociais.

desenho apropriado para a intervenção, da identificação- e resolução- dos problemas recorrentes de implementação operacional que os programas enfrentam.

Dia após dia, lá na ponta onde os serviços são prestados, a realidade teima em revelar sua complexidade, não antecipada no desenho original da intervenção, exigindo soluções *ad hoc* emergenciais. Esses remendos acabam se tornando soluções “provisórias-permanentes” a menos que uma reflexão mais substantiva seja realizada, tão melhor quanto mais bem esclarecida por alguma estratégia de avaliação. Dia após dia, a efetividade social dos programas pode estar sendo comprometida pelas deficiências de cobertura e qualidade dos serviços e produtos entregues, decorrência dos problemas típicos e recorrentes de implementação em contextos complexos como os presentes no país, como a descontinuidade no fluxo de repasses ou restrições de financiamento; dificuldades de articulação federativa ou de relacionamento com outros agentes envolvidos; gargalos na alocação de pessoal em volume suficiente, com a capacidade técnica desejada, em condições de trabalho adequadas, com recursos técnicos e edificações apropriadas; deficiências de diálogo com usuários e de controle social.

Diagrama 1: A Implementação no Centro da Avaliação Sistemática



Fonte: Elaboração própria.

Avaliações sistemáticas e expeditas sobre esses aspectos (trilha A no Diagrama), realizadas com diferentes agentes e atores envolvidos (operadores na ponta, técnicos ao “nível da rua”, usuários,

beneficiários, gestores etc), produzem informações que permitem perceber as lacunas e equívocos do desenho inicial da intervenção e identificar novos problemas que a realidade concreta e dinâmica vai colocando aos operadores dos programas. Em ciclos recursivos de avaliações, com a análise crítica dos achados dessas pesquisas e também dos indicadores de monitoramento, pode-se orientar a equipe do programa a fazer redesenho de processos-chave, redimensionamento de recursos, realocação de pessoal ou introduzir atividades como capacitação, que poderão promover a inovação incremental da intervenção.

Em algumas situações, em que as tentativas de correção dos problemas operacionais diagnosticados pelas avaliações parciais relevam-se soluções paliativas e frágeis, que não levam ao aperfeiçoamento de fato do programa, talvez seja o caso de reconhecer que se está diante de uma problemática maior: a inadequação do desenho do programa, de suas rotinas e/ou agentes implementadores. Talvez não seja necessário demonstrar que o programa não produz os resultados almejados e avançar mais rapidamente para discussão de um novo arranjo operativo para o programa (Trilha B), trazendo o conhecimento estruturado a partir da revisão dos achados das avaliações de implementação realizadas e o conhecimento tácito acumulados pelos gestores e técnicos na operação cotidiana das intervenções.

Em um dado momento, resolvidos ou mitigados os gargalos mais críticos na implementação do programa, coloca-se, de fato, a necessidade de avaliações somativas (trilha C). Este é o momento crítico de apreciação de resultados e impactos, mais diretos ou indiretos, mais tangíveis ou não, decorrentes do programa. Trata-se de avaliar a eficácia e efetividade da intervenção, do quanto ela atendeu a demanda social e política que motivou sua proposição, e de dimensionar seus outros efeitos – positivos e negativos- não antecipados quando do seu desenho. Tal avaliação, expressa por um conjunto amplo de evidências qualitativas e indicadores quantitativos, permitem aos decisores políticos apreciarem o mérito da política ou programa e, apoiados em considerações políticas e outros critérios imanentes às políticas públicas, optar pela continuidade da intervenção, sua modificação, expansão ou descontinuidade.

Formuladores e gestores são, então, eventualmente instados a promover mudanças pontuais, modificações mais expressivas visando sua expansão ou inovações mais radicais no desenho operacional para uma intervenção mais qualificada (trilha D). A demanda social pode ter mudado marginal ou substantivamente: se antes a questão colocada na agenda de políticas públicas era, prioritariamente, por exemplo, garantir acesso universal à escola, eliminar a fome e/ou prover o saneamento básico, a nova problemática social pode estar

requerendo maior ênfase na melhoria da qualidade da educação, alimentação suficiente e balanceada e provimento mais amplo de serviços de saúde. A escala da intervenção pode ter se ampliado, com a incorporação de novos públicos-alvo ou regiões não previstas anteriormente. Em geral, demandas antigas convivem com novas necessidades, em um mesmo território, com públicos-alvo distintos. Um novo ciclo de implementação de programas se inicia, requerendo inovações incrementais para o novo desenho operacional proposto, a serem viabilizados por avaliações de processo; mais à frente, virão as avaliações somativas, que permitirão apreciações sobre a efetividade da nova intervenção e o ciclo se completa novamente...

5. À Guisa de Conclusão: a necessidade de critérios abrangentes na Avaliação

Em qualquer das trilhas, em todas as etapas, da incorporação da Demanda Social como Políticas Pública, do desenho de programas associados e nas avaliações daí decorrentes, na implementação ou na apreciação somativa, não se deve esquecer dos valores e princípios fundantes e legitimadores das intervenções, específicos para cada sociedade e momento histórico. No Brasil, valores republicanos e princípios de cidadania social estão explicitados na Constituição de 1988 e por um conjunto amplo de institutos complementares como leis, estatutos, convenções internacionais, pactos interfederativos, planos nacionais etc. Assim, na avaliação de políticas e programas no Brasil não se pode deixar de considerar os critérios imanentes que legitimaram o atendimento da demanda pública e do desenho da intervenção proposta tais como universalidade de acesso, equidade de oferta, reparação e/ou respeito à diversidade identitária-cultural, sustentabilidade ambiental, entre tantos outros princípios mais específicos como economicidade e probidade administrativa.

Não cabe aqui uma discussão mais ampla de valores, critérios e parâmetros de avaliação de mérito de programas públicos e projetos sociais. Essa é certamente uma discussão necessária no Brasil, já tendo sido realizado em outras sociedades. Nos EUA um número grande de pesquisadores no campo travou esse debate há várias décadas. À época, como comentado anteriormente, os investimentos públicos em programas de combate à pobreza, promoção da qualidade da educação, construção de moradias, qualificação profissional e revitalização urbana requeriam a definição de um conjunto padrão de critérios que distintas equipes de avaliadores, voltados à análise de diferentes intervenções e públicos-alvo, pudessem partilhar e tornar

as avaliações mais comensuráveis e, em alguma medida, comparáveis na estreita margem de intersecção de problemas e contextos comuns. Como bem mostra Bauer (2011), o debate sobre critérios de mérito na avaliação iniciado nos anos 1960 - com Michael Scriven, Donald Campbell e Daniel Stufflebeam- ganhou força nas décadas seguintes - com Egon Guba e Yvonna Lincoln, Carol Weiss, Thomas Schawndt- culminando com sistematização de protocolos por associações técnico-científicas-profissionais de avaliação dos EUA, Europa e, mais recentemente, na América Latina⁶.

De uma perspectiva pragmática, entre as diversas propostas de parâmetros valorativos disponíveis, os critérios de avaliação definidos pelo comitê de avaliadores da Organização de Cooperação para o Desenvolvimento Econômico (DAC/OCDE) há mais de 25 anos atrás parecem constituir a principal referência para avaliação de mérito de políticas, programas públicos, projetos sociais e projetos de desenvolvimento pelo mundo afora. A concisão, universalidade e adaptabilidade da proposta conceitual e sua adoção pelas agências especializadas nas Nações Unidas certamente contribuíram para isso. Não é uma proposta tão minimalista como as adotadas por agências internacionais de fomento - centrada em parâmetros economicistas- nem tão exaustivas como as advogadas pela Associação Americana de Avaliação⁷. Abarca valores para julgamento de mérito de programas públicos e projetos sociais legitimados pela experiência histórica da construção do Estado de Bem-Estar e das Políticas Públicas nos países centrais. Essa universalidade de princípios republicanos tem se mostrado resiliente e também adaptável em diferentes contextos de avaliação, sejam eles de projetos de cooperação internacional, políticas nacionais, programas públicos setoriais ou projetos sociais de pequena escala.

São cinco os critérios de avaliação “DAC/OCDE”, abordando o mérito de programas e projetos quanto à relevância, eficácia, eficiência, sustentabilidade e impacto (PNUD 2009). De forma simplificada, Relevância de um programa ou projeto está associado ao grau de pertinência do mesmo às demandas públicas prioritárias, isto é, sua aderência à agenda de prioridades políticas de uma determinada sociedade. Eficácia de um programa ou projeto é um atributo relacionado ao grau de cumprimento dos objetivos do mesmo ou de atendimento da demanda motivadora do programa. Eficiência, por outro lado, está associado à qualidade e preocupação de como os recursos são utilizados na produção dos resultados do programa. A Sustentabilidade refere-se à capacidade do programa ou projeto em gerar mudanças permanentes na realidade em que atuaram. Por fim, o Impacto diz respeito aos efeitos de médio e longo prazo sobre os beneficiários e sociedade - positivos ou não- direta ou

⁶ Vide, nesse sentido, o documento compilado por Rodriguez Billela et al (2016), consolidando a discussão sobre Dimensões e Diretrizes para Avaliação para a América Latina e o Caribe desde 2014 no âmbito da Rede Latino-americana de Avaliação, em que são apresentados 30 princípios e/ou critérios de avaliação.

⁷ Vide, nesse sentido, os 17 critérios apresentados em https://www.wmich.edu/sites/default/files/attachments/u350/2014/values_criteria.pdf

indiretamente atribuíveis ao programa ou projeto. Em geral, quando considerados os impactos positivos, emprega-se alternativamente o termo Efetividade.

No Brasil atual, avaliações que podem fazer diferença são aquelas que, do ponto de vista político, estejam referidas aos valores republicanos de uma sociedade mais justa, solidária e coesa, idealizados na Constituição de 1988. Do ponto de vista metodológico, avaliações que podem fazer diferença são aquelas que partem de perspectivas mais plurais e estejam voltadas à identificação dos gargalos de implementação dessas intervenções públicas. São essas avaliações que abrirão possibilidades de redesenho de intervenções para, inclusive, integrar mais atividades e mitigar parte da fragmentação com que os programas sociais se encontram. São desses esforços que virão as inovações necessárias à ampliação da efetividade social das políticas e programas.

Bibliografia

BAMBERGUER, M.; SEGONE, M.; TATEOSSIAN, F. **Evaluating the Sustainable Development Goals With a “No one left behind” lens through equity-focused and gender-responsive evaluations.** New York: UN Women, 2016.

BAUER, A. **Avaliação de impacto de formação docente em serviço: o programa Letra e Vida.** São Paulo, USP, 2011 (Tese de Doutorado em Educação), capítulo 1.

CAMPELLO, T.; FALCÃO, T.; COSTA, P. V. **O Brasil sem miséria.** Brasília: MDS, 2014.

CEJUDO, G. Discurso y políticas públicas: el enfoque constructivista. In: MERINO, M.; CEJUDO, G. **Problemas, decisiones y soluciones.** México: CFE-CIDE, 2010.

COTTA, T. C. Metodologias de avaliação de programas e projetos sociais: análise de resultados e impacto. **Revista do Serviço Público**, Brasília, 49(2):103-124, 1998

DEMO, P. **Pesquisa e Informação Qualitativa.** Campinas/SP: Papyrus, 2001.

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.

KROPF, S.P.; FERREIRA, L.O. A prática da ciência: uma etnografia no laboratório. **História Ciências Saúde-Manguinhos**, 4, p. 589-597, 1998.

MONTAGNER, P.; MULLER, L. H. Apresentação. **Cadernos de Estudos Desenvolvimento Social em Debate**, Brasília, DF, v. 24, p. 7-17, 2015

MCDAVID, 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.

PATTON, M. Q. **Qualitative evaluation and research methods.** Thousand Oaks, Sage, 1990.

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

PORTER, T. **Thrust in numbers.** New Jersey: PUF, 1996.

PNUD. **Manual de Planificación, Seguimiento y Evaluación de los resultados de Desarrollo.** Nova York, 2009.

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

RODRIGUEZ BILELLA, P.D. et al. **Diretrizes para Avaliação para a América Latina e o Caribe.** Relac, Buenos Aires, 2016.

ROSSI, P. The Iron Law of Evaluation and other Metallic Rules. MILLER, J. LEWIS, M. **Research in Social Problems and Public Policy.** Jai Press, Vol. 4, p. 3-20, 1987.

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.

WEISS,C. Where Politics and Evaluation Research Meet. **Evaluation Practice**, 14 (1): 93-106, 1993.

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

YANES,P. La experiencia de evaluación de programas sociales en la Ciudad de México. ENAP. **Reflexões para Ibero-América: Avaliação de Programas Sociais**. Brasília, 2009, p.79-88.