



DEMOCRACIA E DIREITOS

DIMENSÕES DAS
POLÍTICAS PÚBLICAS
E SOCIAIS

MITOS DO DESENHO QUASI-EXPERIMENTAL NA AVALIAÇÃO DE PROGRAMAS

Paulo de Martino Jannuzzi, Escola Nacional de Ciências
Estatísticas | ENCE, do Instituto Brasileiro de Geografia e Estatística |
IBGE

RESUMO

Embora a tese da natureza contingencial e humana da produção científica e tecnológica já seja de largo conhecimento de pesquisadores com formação nas Ciências Sociais, a vivência na Administração Pública revela que o positivismo comteano está ainda muito entre equipes técnicas envolvidas nos processos de formulação, avaliação e controle de Políticas e Programas Sociais. Tal postura é particularmente intensa entre pesquisadores que advogam que os desenhos experimentais e quasi-experimentais é o “padrão-ouro” de avaliação de políticas e programas, no que são referendados por manuais de órgãos multilaterais de fomento. Valendo-se de referenciais clássicos da Ciência e da Avaliação, este texto procura contrapor-se aos mitos das pretensas objetividades absolutas e superioridade técnica dos desenhos experimentais e *quasi*-experimentais, reestabelecendo uma perspectiva mais plural e socialmente reconhecida que o Campo das Públicas oferece na produção do conhecimento nessa área temática.

Palavras-chave: avaliação; positivismo, pós-positivismo; políticas públicas.

MITOS DO DESENHO QUASI-EXPERIMENTAL NA AVALIAÇÃO DE PROGRAMAS

JANNUZZI, Paulo de Martino

INTRODUÇÃO / APRESENTAÇÃO¹

Objetividade e neutralidade são princípios recorrentemente citados como fundamentais na Avaliação de Políticas e Programas, especialmente em manuais de organizações multilaterais de fomento. Afinal, em uma cultura de políticas públicas focadas em resultados, as recomendações das Avaliações só seriam aceitas por todos os interessados – políticos, gestores, academia, mídia e sociedade – se estivessem respaldados na mais isenta e robusta pesquisa científica. Nessa perspectiva, somente estudos conduzidos segundo os preceitos da “insuspeita” *Evidence-based Policy Making* (Políticas Públicas baseada em evidências)² se conseguiria prover alguma racionalidade técnica para os ineficazes programas sociais existentes nos países do Sul, trazendo a ordem econômica e o progresso material para as sociedades subdesenvolvidas.

Embora a tese da natureza contingencial e humana da produção científica e tecnológica já seja de largo conhecimento de pesquisadores com formação nas Ciências Sociais, a vivência na Administração Pública revela que o positivismo comteano está muito mais presente que se imagina, em especial, entre equipes técnicas envolvidas nos processos de formulação, avaliação e controle de Políticas e Programas Sociais (YANES 2010). A assertiva “dados e fatos falam por si” ainda é partilhada com vigor em comunidades de Monitoramento e Avaliação (M&A) alinhadas à “escola” das Políticas Públicas baseadas em evidências. O campo de estudos em M&A seria uma subdisciplina da Física Social, cabendo aos pesquisadores e agentes dessas escolas a busca de “leis universais” acerca do funcionamento das Políticas Públicas e “soluções iluminadas” de como garantir seu pleno funcionamento.

Desnudar a subjetividade e parcialidade implícitas nas Avaliações de Políticas e Programas é um dos objetivos desse texto. Por mais que isso possa soar desconfortável para algumas comunidades epistêmicas e de práticas no campo de M&A, não há produção tecnocientífica neutra e infalível, nem método e técnica com maior cientificidade que outras no desenvolvimento das atividades de monitoramento e avaliação de programas. Contrapor-se à ditadura dos desenhos experimentais e *quasi*-experimentais na avaliação é outro objetivo desse texto. Ao contrário do que se certas comunidades epistêmicas advogam, não existe um método padrão-ouro de avaliação de Políticas

¹ Este trabalho é derivado do projeto de pesquisa PQ/CNPq “Políticas Públicas, Mudança Social e Dinâmica Demográfica no Brasil de 1992 a 2014”.

² *Evidence-based Policy Making* designa um conjunto de práticas e conhecimentos relacionados ao ideário da Gestão por Resultados propugnado pela Nova Administração Pública, com fortes implicações no campo de M&A, como discutido mais à frente nesse texto, a partir do manual de Avaliação de Impacto disseminado pelo Banco Mundial (GERTLER *et al* 2015).

Públicas, como não existe um método universal na Ciência. São essas duas questões – a discussão sobre as pretensas objetividades absoluta e superioridade técnica dos desenhos experimentais e *quasi*-experimentais que estruturam as seções seguintes deste texto.

1. A pretensa objetividade absoluta dos métodos e técnicas das Avaliações

Como toda área de conhecimento técnico-científico em que há volume significativo de recursos para pesquisa e potencial prestígio político para as comunidades de pesquisadores praticantes, Monitoramento e Avaliação de Políticas e Programas Públicos constitui um campo de estudos aplicados com crescente produção e de intensa disputa técnica no país e no exterior. Lá fora e aqui dentro, os recursos de órgãos públicos e de agências multilaterais tem viabilizado a formação, manutenção ou expansão de equipes técnicas em universidades, centros de pesquisas e organizações não governamentais dedicadas ao tema, com produção de estudos avaliativos e instrumentos de monitoramento sobre variados temas, com repercussão menor ou maior no setor público, mídia e sociedade.

A julgar pela qualidade técnica e respeitabilidade institucional das equipes e centros de pesquisa envolvidos, muito do que se produz é certamente consistente do ponto de vista técnico e útil, direta ou indiretamente, para análise de Políticas Públicas e para a gestão de programas sociais. Ademais, a robustez metodológica e a relevância substantiva dessa produção são atestadas pelo fato de que boa parte dela é apresentada em seminários de pesquisadores, debatida em oficinas técnicas com gestores de políticas e programas e também veiculada em publicações técnico-científicas, com avaliação externa de especialistas.

Mas é preciso apontar que nem tudo o que se produz, se apresenta e se publica, contudo, mesmo por equipes e instituições experientes resistiria a uma análise do mérito técnico de sua apropriação, para aprimoramento do desenho, gestão ou avaliação de políticas e programas, propriedade certamente desejável em qualquer produto ou relatório no campo aplicado de M&A. As disfunções e distorções desse acervo menos meritório da pesquisa aplicada na área se revelam por diferentes sintomas. Há trabalhos que, embora cumpram os requisitos técnicos necessários para sua legitimação científica (nas comunidades epistêmicas em que são produzidos), não parecem responder perguntas relevantes, pela ambiguidade ou generalidade dos achados, ou ainda pelos recortes a que a questão e o universo de análise vieram a sofrer para se encaixar na técnica. Na área, há certa mistificação na elaboração de parte destes estudos, que vai da fetichização da técnica à mitificação meta-científica ou meta-política dos mesmos, em detrimento da transparência metodológica, lisura ética e responsabilidade republicana que avaliações de programas – mais ainda que trabalhos acadêmicos – devem dispor. Em alguns estudos, os indicadores tomam lugar dos conceitos que os originaram, como por exemplo, assumir como desenvolvimento humano o que a medida – Índice de Desenvolvimento Humano- revela, não o

que o substantivamente se propõe como conceito (IEO 2016)³. O sociometrista empiricista, isto é, a construção de indicadores pela combinação de outras medidas sem um claro marco conceitual ou normativo ganha autonomia como campo de pesquisa aplicada, oferecendo mensurações – com várias casas decimais – para qualquer que seja a dimensão analítica que se queira apreender. Em outros casos, a compulsão pela medida “mais precisa possível” do que se imagina ser o impacto de programas, deixa em segundo ou terceiro planos os esforços de obtenção de indicadores de contexto e processo que podem ajudar a entender o sucesso e insucesso dos programas⁴.

Há estudos em que parece que a escolha do objeto de análise – da política ou programa – é ditada pela possibilidade de aplicação do método, não a relevância social da questão ou dos problemas elencados pelos gestores e técnicos para serem respondidas. A impressão que se tem em boa parte desses trabalhos é que a suposta clareza e “beleza estética” da formulação matemática ou do desenho da pesquisa de campo orienta a escolha do objeto a ser investigado e não o contrário. A compulsão por encontrar relações de causalidade e atribuição acaba tornando-se mais importante que investigar resultados junto a segmentos mais amplos da população beneficiária das políticas ou responder as perguntas que técnicos e gestores tem sobre as dificuldades de implementação dos programas. O método determina o que se quer responder, qualquer que seja a natureza da intervenção (MORAL-ARCE 2014)⁵. Há ainda situações em que determinadas comunidades disciplinares se arvoram de detentoras dos únicos métodos legítimos de avaliação de programas, como a dos “randomistas” como bem os classificou o “insuspeito” Ravallion (2009)⁶. Como explicitado pelos autores de um dos manuais clássicos nessa linha:

As avaliações de impacto fazem parte de uma agenda mais ampla: a da formulação de políticas com base em evidências. Essa crescente tendência global está marcada por uma mudança no enfoque, de insumos para resultados (...)
O monitoramento e a avaliação estão no cerne da formulação de políticas com base em evidências. Fornecem um núcleo básico de ferramentas que as partes

³ Nesse relatório, produzido por agência de avaliação independente das Nações Unidas, há diversas críticas aos Relatórios de Desenvolvimento Humanos produzidos desde 1990 e, em particular, a “autonomização” do Índice de Desenvolvimento Humano e sua “centralidade” nas discussões suscitadas pela divulgação anual dos relatórios, muitas vezes empobrecendo o debate sobre o que é, de fato, relevante para o Desenvolvimento Social, Crescimento Econômico e Proteção Ambiental (a concepção mais atual de Desenvolvimento pelas agências das Nações Unidas).

⁴ Isso é claro no campo educacional. A criação, pelo INEP, dos testes de avaliação de desempenho do Sistema de Avaliação da Educação Básica (SAEB) e a introdução posterior do Índice de Desenvolvimento da Educação Básica como medida de avaliação da “performance educacional” acabou provocando, inclusive, uma “febre de avaliação” com criação de testes de desempenho por secretarias estaduais e municipais de educação. Mais recentemente o INEP resgatou a importância de uma visão mais sistêmica do processo educacional, como revela a proposição do Sistema Nacional de Avaliação da Educação Básica (SINAEB), em que medidas de desempenho escolar compõem parte de uma matriz mais ampla e rica de indicadores educacionais.

⁵ Nas suas palavras “Por desgracia, muchos interesados, a distintos niveles, creen que se puede (y se debe) realizar evaluaciones de impacto de manera rutinaria a todos los programas. Paradojicamente, esa insistencia en tratar de realizar evaluaciones de impacto de manera sistematica, puede conducir al resultado no deseado de desperdiciar recursos (que son limitados) por tratar de realizar una evaluacion de este tipo.” (MORAL-ARCE, 2014, p.40).

⁶ Martin Ravallion foi economista-chefe do Banco Mundial durante muitos anos, tendo publicado diversos artigos no campo da pobreza. Escreveu o texto “Should randomists rule?” em 2010, fazendo uma crítica às crenças e abusos na utilização dos métodos experimentais na avaliação de programas públicos e na tomada de decisão em políticas. Daí sua aludida “insuspeição”.

interessadas podem usar para verificar e medir a qualidade, eficiência e efetividade das intervenções nas várias etapas de execução ou, em outras palavras, focar em resultados.

(...) Neste sentido, a informação e as evidências se tornam meios para facilitar a conscientização do público e promover a responsabilidade governamental. (...) Em um contexto no qual os formuladores de políticas e a sociedade civil exigem resultados e cobram prestação de contas dos programas públicos, a avaliação de impacto pode oferecer evidências críveis e robustas quanto ao desempenho e, fundamentalmente, quanto a se um programa específico atingiu os resultados desejados. (GERTLER *et al* 2015 ,p.03-4).

A crença de que avaliações de impacto experimental ou suas variações constituem-se no padrão-ouro é reforçada, em um círculo "auto-referenciado" pelos bancos multilaterais de fomento e outras comunidades de financiadores de projetos sociais. Essas instituições, em geral, constituídas por equipes com formação acadêmica marcadamente disciplinar e positivista, com pouco conhecimento de desenho e prática de gestão de programas, reforçam essa lógica perversa professada por essa comunidade epistêmica: só colocam recursos em iniciativas em que o gestor se compromete a seguir a cartilha prévia da avaliação de impacto, qualquer que seja a natureza da intervenção, viabilidade operacional do delineamento ou os princípios éticos a obedecer. É o que La Rovere (2014) discute, no contexto de avaliação de políticas ambientais, em que a investigação de contribuições marginais de iniciativa na área e a separação de unidades investigadas em amostra de tratamento e de controle são operacionalmente inviáveis. E desnuda como funciona o círculo de financiamento-método-financiamento de projetos e programas:

Yet pressure arising from multiple sources (donors and evaluation fora) towards the perceived higher rigour achievable through quantitative approaches and attribution is being reapplied on impact assessment and evaluation practitioners. This demand is stimulated (or often enforced) by major donors insisting that a quantitative approach is the only credible one

These influential donors are almost always located in the same places (i.e. countries, cities and often intellectual circles) as the academic institutions where such tools are being promoted (LA ROVERE 2014,p. 285).

Se não há como discordar quanto aos propósitos almejados dos randomistas em contribuir para o aprimoramento do gasto público, há que refutá-los acerca da presunção da avaliação de impacto como "padrão-ouro" de Avaliação de Programas. Sob o mantra da *Evidence-based Policy Making*, a avaliação de programas reduz-se muitas vezes a exercícios aplicados de econometria, em que variações de um mesmo modelo-padrão são "rodados", sem efetiva verificação de pressupostos desejáveis na implementação dos programas ou quanto às propriedades necessárias dos dados empíricos. Supõe-se que os resultados de programas são univariados, quando se sabe que, no desenho dos mesmos, procura-se produzir múltiplos efeitos, frente à diversidade de públicos a atender. Com amostras selecionadas que potencializam a validade interna do desenho metodológico da pesquisa avaliativa (em geral, experimental ou *quasi*-experimental), em detrimento de sua validade externa do público-alvo dos programas e sua representatividade na realidade dura e concreta da implementação de programas em ambientes complexos, geram-se relatórios que, em uma remissão a visões já

ultrapassadas acerca da produção do conhecimento científico, se autodeclaram como “politicamente neutros” e “cientificamente atestados”. Esquece-se que a atribuição (ou deslegitimação) dos efeitos identificados em uma população aos componentes de um programa depende de muitas escolhas quanto aos testes estatísticos, níveis de significância, características e tamanho de amostras, dos pressupostos com relação às propriedades de distribuição dos dados. Não se encontra em muitos desses trabalhos a discussão sobre poder estatístico dos testes usados ou sobre a análise de resíduos após a estimação de parâmetros de modelos. Menos ainda comuns são análises mais exaustivas sobre os potenciais vieses introduzidos na estimação do sentido e intensidade do impacto (ou não impacto) pelas calibrações dos grupos tratamento e controle pela técnica “*propensity score matching*”.

Com todas essas limitações, advogam que Política Pública “boa” e “científica” é a produzida nos laboratórios de econometria, longe das “escolhas subjetivas” dos gestores ou “pressões suspeitas” dos públicos potencialmente favorecidos ou excluídos da “benesse governamental”. Tal como um medicamento, Política Pública precisaria passar pelo ensaio epidemiológico típico de caso-controle de laboratório (de econometria, vale acrescentar), crença que segmentos crescentes da comunidade epistêmica de Saúde Pública, envolvida na gestão de programas, deixou de partilhar, pelos insucessos em garantir as mesmas condições contextuais nos grupos de beneficiários e não beneficiários, ou ainda pela complexidade operacional de programas na área e pelas inevitáveis e incontroláveis sinergias advindas de outras políticas ou dos efeitos do contexto socioeconômico, que se manifestam desigualmente pelo território⁷.

Ademais, os randomistas partilham de uma visão ingênua sobre a utilização de avaliações na modificação ou descontinuidade de políticas e programas, como se os resultados dos estudos fossem automaticamente “comprados” e implementáveis rapidamente pelo decisor público. Desconhecem a complexidade do processo de tomada de decisão no setor público e, sobretudo, a engenharia institucional envolvida na operação dos programas. Como bem lembra Weiss em seu livro clássico *Evaluation*:

Program and policy decisions in a democracy do not take place in an autocratic context. There is rarely a single decision maker who can adopt evaluation results and hand down orders.

(...)

Even when members of the program and policy communities become convinced by the findings and want to take corrective action, they have to generate support in other constituencies, amass resources, allay suspicions, employ staff with appropriate skills, and see that front-line practitioners endorse and implement the changes. (WEISS 1998, p.324)

Curiosamente, essas limitações metodológicas não enfraquecem a crença dos randomistas quanto à robustez dos achados de suas pesquisas e à superioridade das técnicas que empregam frente a de outras comunidades de avaliadores. Como já advertira Ravallion (2010) em seu curto ensaio crítico aos

⁷ Uma análise mais extensa sobre os “mitos” da avaliação experimental pode ser consultada em Ravallion (2009), Moral-Arce (2014) e Jannuzzi (2016a).

randomistas:

From the point of view of development policy-making, the main problem in the randomistas agenda is that they have put their preferred method ahead of the questions that emerge from our knowledge gaps.

(...)

The emphasis that researchers are now giving to obtaining better knowledge about development effectiveness is welcome. Randomization is one of the tools that can help. However, the important task of investigating what works and what does not in the fight against poverty cannot be monopolized by one method. (RAVAILLON 2010, 2-5)

A “ditadura” do método imposto pelas comunidades epistêmicas a seus membros, no processo de formação dos pesquisadores e de legitimação do conhecimento científico produzido é a grande crítica do clássico – embora pouco lido – livro *Contra o Método* de Feyerabend (2011). Como bem coloca Araújo sobre uma das polêmicas suscitadas em seu livro:

(...) em ciência não deve haver um método rígido a ser imposto, eis que seria um entrave para inúmeras descobertas e avanços científicos. A ciência requer pluralidade de procedimentos metodológicos para lidar com a pluralidade de problemas e situações que a afetam. Para que a ciência progrida, deve ter à sua disposição quantos métodos e procedimentos forem necessários para serem escolhidos e empregados em cada situação específica. (ARAÚJO, 2012, p.139)

Criatividade e irreverência foram cruciais para o avanço da Ciência Moderna, como revelou a história da Física de Galileu a Einstein, para citar um campo disciplinar. Nas palavras do próprio Feyerabend, ele adverte:

Especialistas, ou ignorantes que tenham adquirido as insígnias formais de uma especialidade, sempre tentaram e com frequência tiveram êxito nisso: assegurar para si mesmos direitos exclusivos em domínios especiais. Qualquer crítica da rigidez da Igreja Católica Romana [a Galileu e outros iluministas] também aplica-se a seus sucessores modernos, tanto científicos quanto ligados à ciência”. (FEYERABEND 2011, p.175).

2. A ditadura do Quantificacionismo na Avaliação

Essas tendências de empiricismo e reificação das técnicas quantitativas não são exclusivas na pesquisa em M&A, mas se replica em comunidades epistêmicas específicas nas Ciências Sociais Aplicadas. Na Economia esse é um dos diversos pontos de divergência entre “ortodoxos” e “heterodoxos”, como bem analisam Bruno e Caffé (2017). Segundo os autores, os economistas neoclássicos compartilhavam uma noção “fiscalista” dos fenômenos econômicos, que iria muito além do empréstimo de conceitos caros à Mecânica Newtoniana como equilíbrio, inércia, histerese, mas a uma interpretação “naturalista” do comportamento humano, muito alinhada à interpretação de Comte no século XIX acerca das Ciências Sociais.

(...) entendo por Física Social, a ciência que tem o objeto o estudo dos fenômenos sociais considerados dentro do mesmo espírito dos fenômenos astronômicos, físicos,

químicos e fisiológicos, quer dizer, como sujeitos às leis naturais invariáveis, cuja descoberta é o objetivo específico de suas pesquisas (COMTE apud BRUNO e CAFFE 2017, p.28)

Dessa concepção e convicção acerca do funcionamento da Economia e Sociedade é fácil entender porque os neoclássicos defendem com tanta veemência o método experimental ou observacional, e o consideram como o mais legítimo, neutro, imparcial, livre de conteúdos ideológicos na produção de conhecimento científico na Economia ou qualquer outro campo⁸.

Nesse contexto de positivismo – assumido explícita ou ingenuamente – o emprego da técnica pela técnica parece estar tomando o lugar, o tempo e o empenho das equipes, em detrimento de análises multidisciplinares e circunstanciadas com que devem ser tratadas as problemáticas complexas que envolvem a produção dos serviços sociais na educação, saúde pública, qualificação profissional ou desenvolvimento social, para citar algumas das áreas mais proeminentes no campo de M&A no país. A pesquisa na área corre o risco de uma hiper-especialização técnica na investigação das problemáticas que afetam as políticas e programas públicos em detrimento de abordagens mais sistêmicas no entendimento dos mesmos. Alguns centros de pesquisa parecem investir em saber cada vez mais sobre cada vez menos, apostando na ortodoxia clássica do recorte disciplinar dos objetos de análise, decisão que pode ser plausível no mundo acadêmico, mas uma escolha equivocada no universo complexo e interdisciplinar das Políticas Públicas.

Políticas Públicas são objetos complexos de análise, com arranjos operativos envolvendo diversas instituições e milhares de agentes, com diferentes mandatos e capacidades de implementação, produzindo efeitos dispersos no tempo e no território, conforme os efeitos menos ou mais intensos de fatores contextuais pouco controláveis pelos operadores das políticas e programas. Assim, abordagens cartesianas clássicas de avaliação, centradas na estimação da contribuição marginal de atividades e componentes das políticas e programas sobre dimensões socioeconômicas almejadas, parecem ser cada vez menos frutíferas. Desenhos experimentais e *quasi*-experimentais de avaliação de Políticas Públicas não só envolvem questões éticas de difícil contorno – como a escolha de quem faz parte do grupo Controle e do Tratamento – como também problemas de operacionalização nada triviais – “isolamento” dos dois grupos ao longo do tempo e garantia de “isonomia” das demais condições ambientais – e questionamentos crescentes na especificação dos modelos de análise – arbitrariedade das escolhas técnicas de *propensity score matching*, dentre tantos outros. Ou seja, tais avaliações têm dificuldades crescentes na sua proposição, desenho, implementação e análise (JANNUZZI 2011, 2016).

Atribuição de impacto a um programa ou a um componente dele é cada vez mais difícil em contextos

⁸ O primado das “estatísticas neutras” ou das “interpretações dos indicadores” na Economia foi tema de um interessante debate no jornal Folha de S.Paulo em 21 e 24 de agosto de 2016 entre, do lado positivista-ortodoxo-liberal, Samuel Pessoa e Marcos Lisboa, e do outro, plural-heterodoxo-estruturalista, Elias Jabbour e Luis Fernando de Paula.

em que políticas públicas são desenhadas para atender múltiplos objetivos (ainda que com ênfase diferenciada entre os mesmos). Política Pública não é projeto social de abrangência limitada, poucos operadores e públicos restritos, com contextos “controlados”. Como bem registra Bamberguer (*et al.*, 2016) em documento sobre diretrizes de avaliação da Agenda dos Objetivos de Desenvolvimento Sustentável para a ONU Mulheres:

In small projects with a low level of programme complexity, relatively simple institutional arrangements and a low level of contextual dependence, it is possible to trace and evaluate a direct causal relationship between a programme intervention (e.g., drinking water, scholarship for girls to attend secondary school) and the intended outcome (e.g., lower rates of diarrhea, higher rates of girl’s enrolment). As programmes become more complex in terms of the three previous dimensions, the number of inputs increases (often operating differently in different communities or regions), the number of intended and unintended outcomes also increases, and the influence of different stakeholders and institutional arrangements becomes more complicated, as well as the number of contextual factors. Consequently, it becomes increasingly difficult, or in many cases impossible, to determine direct causal relationships (BAMBERGUER *et al.* 2016:46- 47)

Paradigmas e modelos de médio alcance relativamente seguros para garantir produção acadêmica que, por sua vez, gerem resultados que viabilizem premiações, novos financiamentos para pesquisas e formação de quadros é um mecanismo clássico de funcionamento da Ciência contemporânea, de forma mais explícita e estruturada do que Khun (1992) ou que os pesquisadores dos Estudos Sociais da Ciência poderiam imaginar⁹. A diferença – e o problema – é que na área de M&A, como bem observou Worthern (*et al.*, 2000) em situações assemelhadas e concretas nas políticas educacionais nos Estados Unidos nos anos 1960, essa produção científica com viés sociométrico e/ou economicista acaba muitas vezes sendo aceita de forma pouco crítica ou voluntarista na Administração Pública e, pior, orientando decisões cruciais acerca de mérito, desenho e gestão de programas públicos, mesmo quando a consistência dos resultados são questionáveis¹⁰.

Como bem assinalou Porter (1995), o “quantificacionismo” é um instrumento poderoso para tomada de decisão aparentemente imparcial e justa em Políticas Públicas, conferindo autoridade a quem não a conquistou legitimamente pelo voto ou por delegação. Para certas comunidades tecnocráticas e plutocráticas na Administração Pública, as decisões baseadas em dados, indicadores e evidências produzidas em avaliações quantitativas e protocolos estruturados gozam de neutralidade e objetividade incontestáveis. Na realidade, dados e indicadores resultam de uma série de escolhas políticas (sobre que aspectos observar e medir) e preferências metodológicas (sobre como observá-los e medi- los).

⁹ Essas visões idealísticas e positivistas da Ciência partilhadas na comunidade de pesquisadores das Ciências Sociais Aplicadas, depois de mais de cinquenta anos da “denúncia” khuniana da natureza humana e idiossincrática da produção do conhecimento científico e do acervo dos Estudos Sociais da Ciência é revelador de um problema muito grave na ementa dos cursos de Epistemologia ou Metodologia da Ciência nos cursos de graduação e pós-graduação.

¹⁰ No Brasil, o Pronatec – Programa Nacional de Acesso ao Ensino Técnico e ao Emprego – operado entre 2011 e 2015, é um bom exemplo de como um programa com uma série de méritos (JANNUZZI 2016a) pode ser descontinuado por avaliações instrumentalizadas por esse viés sociométrico-economicista-liberal (BARBOSA Fo. et al 2016).

Nas palavras do autor:

The capacity to yield predictions or policy recommendations that seem to be vindicated by subsequent experience doubtless counts in favor of a method or procedure, but quantitative estimates sometimes are given considerable weight even when nobody defends their validity with real conviction. The appeal of numbers is specially compelling to bureaucratic officials who lack the mandate of popular election, or divine right. (...) A decision made by the numbers (or by explicit rules of some sort) has at least the appearance of being fair and impersonal (...). Objectivity lends authority to officials who have very little of their own. (PORTER, 1995:p.8).

Como ilustra Latour (2004), já são conhecidas, de longa data, as estratégias retóricas de sustentação de procedimentos e achados na produção científica nos laboratórios e centros de pesquisa¹¹, entre elas o “quantificacionismo”. De fato, a leitura de várias avaliações e estudos aplicados à análise de políticas no país revela algumas delas: a busca do argumento de autoridade pela citação de bibliografia clássica ou, cada vez mais, de pesquisadores e seus orientados de um centro ou universidade com forte financiamento externo; a matematização dos procedimentos e a dedução lógica dos resultados; a discricionariedade na escolha de amostras, construção de indicadores, tipos de testes de hipóteses e de níveis de significância; e por fim, mas não menos frequente, a adoção de hipóteses “heroicas” e/ou supostos *ceteris paribus* no começo da pesquisa (e explicitados no meio do relatório final), mas esquecidos quando da apresentação das conclusões.

Aos olhos de pesquisadores de outras comunidades epistêmicas mais plurais no campo de M&A, a “certeza”, a “imparcialidade” e “meta-cientificidade” com que são apresentados as avaliações de políticas e os programas produzidos nessa abordagem “quantificacionista” não deixam dúvidas quanto à natureza mítica e dogmática a preceitos positivistas acerca da produção do conhecimento. Mas o campo das Políticas Públicas – e suas sub-áreas como a M&A – não pode ser enquadrado como uma Física Social ou uma Microeconomia do Bem-Estar, investigada somente com o arsenal de técnicas de inspiração positivista. Estas ferramentas são insuficientes e limitantes para se entender a complexidade e vastidão em que o campo das Políticas Públicas vem se transformando, um século depois de crescimento sistemático enquanto área de intervenção estatal¹² e muitos outros enquanto Filosofia Política. Como outras disciplinas científicas que aqui no Brasil demarcam o campo das Ciências Sociais Aplicadas, a produção de conhecimento no “Campo das Públicas” requer, para além das tradições nomológicas-explicativas inspiradas nas Ciências Naturais, a hermenêutica histórico-

¹¹Estratégias retóricas que, inclusive, são empregadas nesse texto como a preferência por citações literais em detrimento de paráfrases, para evidenciar a corroboração de diversos autores às discussões aqui apresentadas. O uso dos anglicismos *Evidence-based*, *Rules-dictated* e *Equity-guide* também poderia ser enquadrado nessa estratégia.

¹² Até início do século XX, os gastos com financiamento das atividades estatais mantinham-se por volta de 7% do Produto Interno Bruto nos diferentes países europeus. Com o ativismo operário, guerras e a instabilidade econômica os Estados passaram a ter um papel não apenas regulador, mas intervencionista no campo econômico e social. Assim, nos anos dourados do capitalismo, no pós-guerra, a necessidade de financiamento dos Estados de países europeus para promoção das Políticas Públicas (na forma de isenções ou programas) já estava além dos 30%, cifra que o Brasil só chegou e se estabilizou nos anos 2000.

interpretativa das Ciências Humanas¹³.

Ademais, o *modus operandi* da Ciência Moderna, produzida em contexto acadêmico, disciplinar, hierárquico e estável já não é funcional em muitas áreas contemporâneas de conhecimento, muito menos no “Campo das Públicas”. Tal como outras áreas dinâmicas de produção de conhecimento, ela se vem se reestruturando em bases institucionais e epistemológicas mais amplas como as que Pombo (2006) denomina de Interciências¹⁴. Não basta “emprestar” conceitos, categorias de análise e paradigmas de outras ciências; é preciso “re-significá-los” frente aos objetos de análise em questão; é preciso integrá-los na interpretação dos fenômenos e não apenas agregar diferentes perspectivas de análise disciplinar.

Talvez mais do que uma “Interciência”, o “Campo das Públicas” seria, na terminologia emprestada de Echeverría (2009), um ramo exemplar da Tecnociência, com produção de conhecimento fora do laboratório clássico, em contexto transdisciplinar, heterogêneo, heterárquico e instável. Se na Ciência Moderna o conhecimento “válido” precisa ser produzido em laboratório e centros de pesquisa com algum grau de “acreditação” anterior e necessita passar por um ritual controlado de legitimação, assegurado por “pares” acadêmicos em programas de pós-graduação ou em publicações especializadas, na Tecnociência, os estudos, as técnicas e as inovações são gerados – e avaliados – por um número muito maior de agentes, além dos acadêmicos, em circuitos mais plurais e ritos mais descentralizados que os convencionais. Se a Ciência basta a si mesmo em seus conceitos e explicações, a Tecnociência precisa ultrapassá-los em direção à aplicação. O campo de conhecimentos conexos às Políticas Públicas está, pois, mais ligado ao contexto plural e eclético da Tecnociência do que o cartesianismo da Ciência Moderna.

Na Ciência, Interciência ou Tecnociência não existe objetividade absoluta, mas objetividade socialmente construída, como propõe Fourez (1995). Objetividade construída e partilhada pelas comunidades epistêmicas específicas, que demarcam o que é ou não é “científico”, o que deve ou não ser observado ou estudado. Nas suas palavras:

Quando se está habituado a ver o mundo de certo modo, torna-se quase impossível ver as coisas de maneira diferente. Questionar essa visão criaria uma profunda crise afetiva. A visão que se tem do mundo surge então como absolutamente objetiva (FOUREZ 1995, p.55)

Enfim, a Ciência, Interciência ou Tecnociência – e Avaliações – tem partido. É o partido, valores e visões da comunidade científica. E como bem coloca Fourez (1995), a comunidade científica pertence à classe média nas sociedades contemporâneas, identificando-se com a ordem social existente e

¹³ Estas são as duas categorias ou tradições de pesquisa que Habermas (2009) insiste em pontuar, a despeito da negação do cientificismo da última pelos positivistas.

¹⁴ Para a autora, Ecologia e Neurolinguística seriam exemplos de Interciências. Para além dessas categorias em que o conhecimento científico contemporâneo se desenvolve mais rapidamente, ela propõe ainda as Interdisciplinas como a Sociologia das Organizações e Pesquisa Operacional e as já mais conhecidas Ciências de Fronteira, como a Geografia Econômica e Engenharia Genética (POMBO 2006).

disputando recursos com outros segmentos e setores da sociedade. Mas reconhecer a natureza humana e idiossincrática da produção de conhecimento não o fragiliza. O que o enfraquece é a falta de transparência metodológica; é a dogmatização de princípios e procedimentos não consensuais; é a tentativa de se apresentarem como neutros e insuspeitos a qualquer prova os resultados de pesquisas quando se sabe que diversas decisões técnicas foram intrinsecamente subjetivas e, de partida, apoiadas em visões e valores de mundo, de Ciência e de Políticas Públicas que não são universais.

CONSIDERAÇÕES FINAIS

Embora os manuais aplicados na área, sobretudo os disseminados por organizações multilaterais de fomento, dediquem-se em apresentar as etapas, métodos e técnicas de monitoramento e avaliação como se fossem universais, neutros, livres de uma concepção ideológica e política fundante, as práticas na área estão longe de assim se revelarem. Evidências de pesquisas e de auditorias revelam “verdades” que derivam de escolhas ou apostas anteriores – explícitas ou não – acerca de valores político-ideológicos e de princípios epistêmicos acerca da produção de conhecimento em Políticas Públicas. Como todo campo de conhecimento, a avaliação de políticas e programas é uma atividade humana, orientada por valores ideológicos e paradigmas diversos, regida por métodos e técnicas validadas por diferentes comunidades epistêmicas, produzindo resultados contingentes, que podem ser confirmados, refutados ou superados, em algum momento, por novas investigações. Ou então, não seriam produção técnica-científica. Falibilidade do conhecimento técnico é antes uma virtude que uma fragilidade; denota antes a natureza científica e não dogmática do mesmo (SACCO 2016).

Encarada como sub-campo da Interciência ou da Tecnociência, a Avaliação é sim uma atividade influenciada por valores humanos, políticos e ideológicos (VAITSMAN & PAES-SOUSA 2009). Se o desenho e a implementação de políticas e programas são construções políticas por que não a Avaliação não o seria? Afinal,

(...) the policies and programs with which evaluation deals are the creatures of political decisions. They were proposed, defined, debated, enacted, and funded through political processes and in implementation they remain subject to pressures – both supportive and hostile – that arise out of the play of politics.

(...)

The programs with which the evaluator deals are not neutral, antiseptic, laboratory-type entities. They emerged from the rough and tumble of political support, opposition and bargaining. (WEISS 1993, p.94)

Avaliação não é uma investigação neutra com respeito aos valores de quem a realiza, seja ele o formulador comprometido com a criação da política; seja um pesquisador externo ou instituição contratada (YANES 2009). A narrativa dos resultados da avaliação depende, como se procurou argumentar nesse texto, dos valores imanentes das Políticas Públicas. Na mesma linha, advoga Uitto (2014), em defesa da Sustentabilidade Ambiental como mais um valor público a ser considerado na avaliação de políticas e programas:

Although one role of evaluation is to generate the most objective and verifiable data and information as possible for decision-making, evaluation is a societal function that must be anchored in the values and goals that we hold. Evaluation should be guided by moral principles and understanding the underlying processes, nor just measuring outcomes (RALLIS *apud* UITTO 2014, p.8).

Assim, no debate sobre impactos ou não das políticas e programas, o que está em disputa é muito mais que a validade interna do método, a suposta superioridade de uma ou outra técnica, a robustez do teste estatístico produzido no laboratório ou a validade externa e riqueza analítica dos achados empíricos em estratégias multi-métodos de investigação. O que está em disputa é muito mais que a objetividade absoluta ou objetividade socialmente construída do conhecimento técnico-científico, contrapondo positivistas e pós-positivistas¹⁵.

Na realidade, o que está em disputa nas avaliações são visões político-ideológicas sobre Estado e natureza das Políticas Públicas, relevadas pelas escolhas de que valores expressam os fins (e os meios) da ação pública: a eficiência do gasto público frente a outros usos que o mercado poderia fazer do mesmo; a eficácia no cumprimento do rito legal dos procedimentos técnico-burocráticos programados frente a escolhas discricionárias socialmente legítimas ou a efetividade social advinda da operação pactuada e legitimada das políticas públicas (JANNUZZI, 2016). Como bem coloca Berrones (2014) em texto que discute “valor público”:

La producción del valor público necessita no solamente reglas institucionales y una eficiente gestión pública, sino también de valores del buen obrar que permiten fortalecer los contenidos, objetivos y metas de las políticas públicas. (BERRONES 2014, p. 60).

Visões de mundo e de papéis do Estado – Estado Liberal, Estado Regulador e Estado do Bem-estar Social – conformam desígnios e arquiteturas diferentes de Políticas Públicas, valorizando distintamente a eficiência econômica, eficácia procedural e a efetividade social como critérios de formulação e avaliação da ação pública. Ainda que todos sejam valores republicanos intrínsecos às sociedades democráticas, a primazia de um sobre os demais está condicionada à concepção subjacente de Justiça Social e a forma de alcançá-la; por sua vez tal concepção aponta o modelo de avaliação a ser empregado que, ao final, traz evidências que reforçam ou não as visões de mundo e as escolhas de políticas públicas realizadas.

Enfim, a seleção de políticas e programas a avaliar, o escopo de questões a serem respondidas, a escolhas metodológicas de como tratá-las não são neutras ou imparciais. Essas decisões são orientadas explicita ou implicitamente pelos valores públicos e motivações políticas compartilhadas pelos grupos de interesse envolvidos no financiamento, controle ou operação das Políticas Públicas. Ancorar-se em valores públicos não é um problema na avaliação; problema mesmo é a dogmatização por que ela

¹⁵ Pelo impreciso rótulo de pós-positivistas está se incluindo um contingente bastante heterogêneo de pesquisadores que partilham da concepção da ciência como atividade socialmente produzida e ritos de validação construídos. Fosse esse texto uma discussão no âmbito dos Estudos Sociais da Ciência essa imprecisão não seria certamente aceitável.

passa em determinadas comunidades epistêmicas.

Referências

- ARAÚJO, P.S. Feyerabend e o pluralismo. In; VIDEIRA, A.A.P. (org). **Perspectivas contemporâneas em filosofia da ciência**. Rio de Janeiro: Ed.Uerj, 2012, p.131-164.
- 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.
- BARBOSA Fo., F. H. *et al.* Pronatec Bolsa-Formação Uma Avaliação Inicial sobre Reinserção no Mercado de Trabalho Formal. **Encontro Nacional da ANPEC**, Florianópolis, 8 a 11 de dezembro, 2015.
- BERRONES, R.U. La importancia de la ética en la formación de valor público. **Estudios Políticos**, 32, pp. 59-81, 2014.
- ECHEVERRÍA, J. Interdisciplinarietà y convergencia tecnocientífica nano-bio-info-cogno. **Sociologías**, Porto Alegre, ano 11, v.22, p.22-53, jul/dez 2009.
- FEYERABEND, P. **Contra o Método**. São Paulo: Ed.Unesp, 2011.
- FOUREZ, G. **A construção das Ciências**. São Paulo: Ed. Unesp, 1995.
- GERTLER, P. *et al.* **Avaliação de impacto na prática**. Washington, Banco Mundial, 2015.
- HABERMAS, J. **A lógica das Ciências Sociais**. Petrópolis: Vozes, 2009.
- HOWLETT, M.; RAMESH, M; PERL, A. **Política pública, seus ciclos e subsistemas**. Rio de Janeiro: Campus, 2013.
- IEO. **Evaluation of the contribution of UNDP Global and regional Human Development Reports to the public policy processes**. New York: PNUD, 2016.
- JANNUZZI, P.M. Avaliação de Programas Sociais no Brasil: Repensando Práticas e Metodologias das Pesquisas Avaliativas. **Planejamento e Políticas Públicas**. , v.36, p.251-275, 2011.
- JANNUZZI, P.M. **Monitoramento e avaliação de programas sociais: uma introdução aos conceitos e técnicas**. Campinas: Alínea, 2016.
- KUHN, T. **A Estrutura da Revoluções Científicas**. São Paulo: Perspectiva, 1992.
- LA ROVERE, R. The consultative group on international agricultural reseaech approach to impact evaluation on environment and natural resources management, pp. 277-288. In UITTO,J.I. (org) **Evaluating environment in international development**. New York: UNEP/Routledge, 2014.
- LATOUR, B. **Ciência em ação: como seguir cientistas e engenheiros sociedade afora**. São Paulo: Unesp, 2000.
- MINAYIO, C. *et al.* **Avaliação por triangulação de métodos: abordagem de programas sociais**. Rio de Janeiro: Ed.Fiocruz, 2005.
- MORAL-ARCE, I. **Elección del método de evaluación cuantitativa de una política pública: buenas prácticas en América Latina y la Unión Europea**. Madrid: EuroSocial, 2014. (Documentos de Trabajo, n. 6).

MORENO, L. *et al.* **Los sistemas de bienestar europeos tras la crisis económica**. Madrid: EuroSocial/IILA, 2014.

POMBO, O. Práticas interdisciplinares. **Sociologias**, Porto Alegre, ano 8, v.15, p.208-249, jan/jul 2006.

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

RAVALLION, M. **Should the randomists rule?** Bekerley Eletronic Press, pp.01-05, fev. 2009.

SACCO, F. O caráter social da ciência e os seus desafios na contemporaneidade. In: ROBERT, P. *et al* (orgs). **Metodologia em Ciências Sociais Hoje**. Jundiaí: Pacco Editorial, 2016, vol.1, p.117-136.

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

UITTO, J.I. Evaluating environment in international development. In _____ (org) **Evaluating environment in international development**. New York: UNEP/Routledge, 2014, p.3-16.

VAITSMAN, J.; PAES-SOUSA, R. Avaliação de programas e transparência da gestão pública. ENAP. **Reflexões para Ibero-América: Avaliação de Programas Sociais**. Brasília, 2009, p.11-23.

WEISS, C. **Evaluation research**. New Jersey: Prentice Hall, 1998.

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: EdUsp/Ed. Gente, 2004.

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