

Ensaio

Sociologia e Avaliação nos EUA: Questões sobre a relação entre metodologia e teoria

Sociology and Evaluation: Issues in the relationship between methodology and theory

Nicoletta Stame^{1*} ¹Sapienza University of Roma, Roma, Italy

Nicoletta Stame, white, emerita professor at Sapienza University of Roma.

COMO CITAR: Stame, Nicoletta. (2023). Sociologia e avaliação nos EUA: Questões sobre a relação entre metodologia e teoria. *Revista Brasileira de Avaliação*, 12(3), e124523. <https://doi.org/10.4322/rbaval202312045>

Resumo

A avaliação baseada em teoria (ABT) enfatiza o pensamento teórico para compreender a dinâmica entre os insumos e resultados de um programa, oferecendo uma alternativa mais matizada às avaliações orientadas por métodos. Ela defende o pluralismo metodológico, garantindo que o método de avaliação esteja alinhado com o problema específico. A ABT é notável por sua profundidade na análise não apenas da ocorrência de resultados, mas também de sua relevância para diferentes grupos, contextos e as razões por trás deles. Essa abordagem evoluiu notavelmente durante a década de 1960 nos EUA, particularmente com os programas da Grande Sociedade, que exigiam avaliações de intervenções sociais abordando questões como degradação urbana e discriminação racial. Figuras-chave como Ann Oakley defendiam a integração de percepções qualitativas em métodos experimentais, marcando uma mudança de foco. O Bureau de Pesquisa Social Aplicada (BASR) influenciou significativamente a ABT com sua ênfase em teorias de alcance médio e métodos de pesquisa integrados, embora seu potencial total em avaliações em larga escala não tenha sido completamente realizado. Este artigo, revisitado para esta publicação a pedido de seu editor-chefe, destaca a capacidade das ABT de lidar com fenômenos sociais complexos e seu foco em compreender tanto os resultados esperados quanto os inesperados de programas, movendo-se além do rigor meramente metodológico para uma compreensão mais ampla da mudança social.

Palavras-chave: Avaliação baseada em teoria. Teoria de programa. Sociologia da avaliação. Políticas sociais.

Abstract

Theory-based evaluation (TBE) emphasizes theoretical thinking to understand the dynamics between a program's inputs and outcomes, offering a nuanced alternative to method-oriented evaluations. It advocates for methodological pluralism, ensuring that the evaluation method aligns with the specific problem. TBE is notable for its depth in analyzing not just the occurrence of outcomes, but also their relevance to different groups, contexts, and reasons behind them. This approach evolved notably during the 1960s in the U.S., particularly with the Great Society programs, which required evaluations of social interventions addressing issues like urban degradation and racial discrimination. Key figures like Ann Oakley advocated for integrating qualitative insights into experimental methods, marking a shift in focus. The Bureau of Applied Social Research (BASR) significantly influenced TBE with its emphasis on middle-range theories and integrated research methods, although its full potential in large-scale evaluations was not completely realized. This article, revisited for this publication at the request of its editor-in-chief, highlighted TBE ability to handle complex social phenomena and its focus on understanding both expected and unexpected program outcomes, moving beyond mere methodological rigor to a broader understanding of social change.

Keywords: Theory-based evaluation. Program theory. Sociology of evaluation. Social policy

A RBAVAL apoia os esforços relativos à visibilidade dos autores negros na produção científica. Assim, nossas publicações solicitam a autodeclaração de cor/etnia dos autores dos textos para tornar visível tal informação nos artigos.

Recebido: Novembro 23, 2023

Aceito: Novembro 27, 2023

***Autor correspondente:**

Nicoletta Stame

E-mail: nstame@gmail.com



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.



Apresentação

Este capítulo foi publicado pela primeira vez em um livro dedicado à relação entre a avaliação e as disciplinas das ciências sociais, "*Mind the gap: Perspectives on Policy Evaluation and the Social Sciences*", editado por Jos Vaessen e Frans Leeuw pela Transactions Publishers, New Brunswick, NJ, em 2010. Ele abordou a relação entre a avaliação e a sociologia, concentrando-se em uma época e local específico dessa última: a sociologia americana no século XX. Essa conjuntura particular no pensamento sociológico, teve uma conexão especial com o debate entre os avaliadores, porque possibilitou contestar a crença dominante na supremacia metodológica, apresentando um impulso para uma ligação saudável entre o pensamento teórico e a metodologia.

O capítulo tem uma tendência histórica, identificando fases de desenvolvimento que correspondem às principais questões surgidas no campo. Como acontece em todas as ciências sociais, duas abordagens principais se confrontaram, mas as falhas não eram claras, resultando em muitas interações inusitadas.

Quando o editor desta revista me pediu para revisitar esse capítulo, fiquei impressionado ao perceber que uma dinâmica semelhante ainda estava em ação: os dois campos principais ainda estavam se confrontando, embora tivesse evoluído cada um à sua maneira, mas as ocasiões para um diálogo não estavam faltando. Portanto, senti-me tentada a acrescentar um apêndice no qual atualizo a narrativa até hoje.

Introdução

As abordagens da avaliação baseada em teoria (do inglês *Theory-based evaluation*¹), que agora desempenham um papel importante na avaliação, enfatizam a necessidade de utilizar o pensamento teórico como uma forma de contornar a armadilha da caixa preta, na qual a incapacidade de uma avaliação de explicar o que acontece entre o insumo do programa e seu resultado leva a resultados insatisfatórios (Stame, 2004). As teorias podem esclarecer não apenas como algo bom deveria acontecer (normativo) ou se algo aconteceu ou não aconteceu (descritivo), mas também para quem, onde e por que aconteceu (explicativo)². Em particular, elas podem explicar por que as coisas aconteceram ou não. Essas teorias podem ser as teorias de mudança subjacentes, muitas vezes implícitas, que os formuladores de políticas, os funcionários da linha de frente e afins têm em mente quando projetam e implementam algum programa (Leeuw, 2003), ou podem ser elaboradas pelas partes interessadas, ou pelo avaliador. Elas podem, portanto, ser diversas, múltiplas e conflitantes, e a tarefa da avaliação é selecionar as que merecem ser investigadas e colocá-las à prova (Weiss, 1997).

Conforme observado por Pawson & Tilley (1997), as abordagens de ABT³ se afirmaram em oposição à profusão de abordagens orientadas por métodos que prevaleceram no passado recente, especialmente entre os avaliadores de programas sociais que receberam uma educação sociológica nos EUA. As abordagens de ABT não são específicas a um método: o método deve se adequar ao problema em questão, e nenhum método é preferível por si só (pluralismo metodológico). Este ensaio examinará o papel da sociologia dos EUA na definição desse cenário. Ele investigará a maneira como a disciplina reagiu ao chamado para a avaliação de programas sociais durante a onda de programas da Great Society na década de 1960, examinará como a tensão entre metodologia e teoria na disciplina influenciou o progresso da avaliação e, por fim, investigará o legado da sociologia norte-americana com relação aos recentes desenvolvimentos na prática da avaliação.

¹ Também referenciada como TBE (inglês) ou ABT (português).

² Sobre a distinção entre os tipos de teorias, consulte Chen & Rossi (1983), Stern (2005), Mark et al. (2000). No entanto, essa distinção pode ser enganosa, especialmente considerando o fato de que as teorias explicativas devem muito às descrições do que, na maioria das vezes, não funcionou (obrigado a Osvaldo Feinstein por sugerir isso em correspondência pessoal).

³ Avaliação Baseada em Teoria.



Meu interesse por essa questão foi recentemente aguçado pelo livro *Experiments in Knowing* (2000), de Ann Oakley. Oakley luta por uma metodologia sólida⁴ que seja capaz de inserir percepções qualitativas na tradição experimental, uma metodologia desenvolvida no início da sociologia americana e utilizada em questões sociais, desde a educação até a criminologia, incluindo a avaliação. De qualquer forma, foi sua concentração na metodologia que me ajudou a enquadrar o que vinha sendo meu principal problema de pesquisa há algum tempo: por que a tradição sociológica proveniente do Bureau of Applied Social Research (BASR) - onde as teorias de “médio alcance” e os métodos de pesquisa estavam tão interligados - não desempenhava o papel central na avaliação para o qual estava tão bem qualificada.

Passei a acreditar que o que importava na avaliação era a capacidade de combinar ferramentas teóricas e metodológicas que pudessem servir de base não apenas para testar a causalidade linear subjacente aos programas concebidos como “ações racionais”, mas também para explicar - para usar a expressão de Merton (1936, p. 894) - “os resultados inesperados da ação social intencional”. Enquanto as tradições sociológicas que enfrentaram o desafio da avaliação de programas sociais consideraram a causalidade linear como certa e se preocuparam principalmente com o método, elas permaneceram incapazes de compreender o cerne do impulso avaliativo. E assim que alternativas começaram a surgir - como estão surgindo agora - a avaliação passou a ser mais capaz de abordar os problemas com os quais se deparava no campo das políticas.

Programas sociais e a origem da sociologia nos EUA

Na década de 1960, os problemas sociais de degradação urbana, desemprego e discriminação racial se tornaram tão difundidos nos EUA que foi declarada uma guerra contra a pobreza, como parte do programa “Great Society” de L.B. Johnson. Os programas públicos foram implementados como uma alternativa aos mercados com mau funcionamento. Isso era tão estranho à tradição americana (a despeito da experiência do New Deal) que foi incluída uma legislação de caducidade, estabelecendo que esses programas, muitas vezes concebidos como experimentos, seriam avaliados periodicamente para que se pudesse decidir se deveriam continuar, mudar ou suspendê-los. Em uma linha otimista, os programas sociais foram concebidos como soluções para os problemas que afetavam os beneficiários, que encontrariam empregos e viveriam melhor, frequentariam escolas melhores e embarcariam em carreiras satisfatórias, teriam casas melhores que facilitariam a vida familiar e assim por diante. Assim, foi dada muita atenção ao comportamento social como resultado, e sociólogos e psicólogos sociais foram encarregados de fazer as primeiras avaliações. Esse vínculo disciplinar resultou em um confinamento às preocupações metodológicas que se tornaram a chave para o sucesso da reforma social.

A sociologia americana foi fundada com base na ideia de que era uma “ciência” (em oposição às artes e às humanidades) e que “a sociedade era um laboratório”, no qual as hipóteses sobre o comportamento social, formuladas principalmente como teorias lineares, do tipo se “a” então “b”, tinham de ser testadas pelo método científico e seus cânones de objetividade, rigor e empirismo (Oakley, 2000). No início do século XX, o interesse em questões metodológicas era predominante sobre as questões teóricas para ambas as escolas concorrentes da sociologia científica: a escola de Columbia, que promovia experimentos sociais, e a escola de Chicago, onde a ênfase estava nos métodos qualitativos de pesquisa (consulte também Coleman (1990, pp. 617-623).

Durante a primeira metade do século, as coisas evoluíram. Em Columbia, Paul Lazarsfeld e Robert Merton colaboraram com o Bureau of Applied Social Research⁵. Eles se concentraram na “sociedade em movimento”, em vez de na “sociedade como laboratório”. Eles direcionaram sua atenção às teorias de uma forma até então desconhecida. Lazarsfeld acrescentou à

⁴ Escrevendo a partir de um ponto de vista de gênero, o impulso de Oakley (2000, parte I) para conciliar métodos qualitativos e quantitativos está preocupado em superar a equação “feminismo é igual à pesquisa qualitativa” e pares semelhantes.

⁵ Utilizei em grande parte o trabalho de Fabrizio Martire (2006) e me beneficieei de sua disposição em compartilhar seu profundo conhecimento sobre o assunto ao desenvolver esta parte do capítulo.



experimentação uma gama maior de estratégias metodológicas que eram mais adequadas aos novos interesses de pesquisa. Merton (1949) desenvolveu uma teoria sobre funções latentes e manifestas como uma crítica ao funcionalismo clássico⁶ e se distanciou da grande teorização de Talcott Parsons⁷, que era muito influente no cenário acadêmico da época, enfatizando a ligação entre a pesquisa e as teorias de “médio alcance” (Merton, 1968). Essas últimas deveriam fazer generalizações de médio alcance a partir da observação das consequências pretendidas ou não da mudança social planejada (consulte também Pawson, 2010).

O ponto de virada na sociologia dos EUA é considerado o *American Soldier*, um estudo maciço sobre as atitudes dos soldados em relação ao combate realizado por Stouffer (1949), um sociólogo de Chicago, com a colaboração de sociólogos e psicólogos sociais de várias escolas. Entre os muitos experimentos realizados, alguns poderiam ser considerados verdadeiros projetos de pesquisa avaliativa, especialmente no caso de estudos sobre os efeitos da propaganda. Além disso, os dados produzidos por essa pesquisa ofereceram a Merton o material para a elaboração de suas famosas teorias de médio alcance do “comportamento de grupos de referência”, que se baseavam nos conceitos de “socialização antecipatória”, “privação relativa” e “lealdades divididas”, como mecanismos que vinculavam o comportamento individual à estrutura social por meio da estrutura de papéis (Merton, 1949; Martire, 2006).

Quando os sociólogos foram chamados a enfrentar o desafio da avaliação de programas sociais (Caro, 1971), tanto a tradição experimentalista quanto a do Bureau pareciam dotadas de um aparato conceitual e de uma metodologia de pesquisa adequados⁸. A tradição do experimento social, que nasceu em Columbia, mas que naquela época também havia migrado para o Meio-Oeste (Universidade de Chicago e Universidade Northwestern), prometia testar se um resultado era o efeito líquido de uma intervenção social. A tradição de *survey*, desenvolvida no Bureau, proporcionou um meio de entender os efeitos das ações sociais planejadas sobre as atitudes e os comportamentos das pessoas. Para ambos, no entanto, o envolvimento metodológico tornou-se o cerne do esforço de avaliação, e as preocupações teóricas que, apesar disso, preocupavam mentes brilhantes em ambos os campos (Campbell e Merton, respectivamente) foram relegadas ao segundo plano. No início, a tradição experimental ocupou o centro do palco; depois disso, suas sortes se alternaram. Elas ainda se confrontam hoje, embora sob diferentes aspectos.

Avaliação experimental

O projeto experimental, na versão dos (quase) experimentos propostos por Campbell, levava em conta a necessidade de realizar experimentos fora do laboratório e no mundo real, e logo passou a ser considerada “a regra de ouro” para a avaliação de programas. Foi creditada a ele a capacidade de avaliar a eficácia de uma intervenção (um tratamento, a variável independente) demonstrando que ela era a “causa necessária” desse efeito (variável dependente), e que seus resultados poderiam ser repetidos em qualquer lugar, desde que fosse administrada da mesma forma (generalização)⁹. Experimentação se tornou a palavra de ordem para a avaliação: não apenas com relação ao design da avaliação, mas também para os programas, concebidos como experimentos sociais cada vez maiores, conforme registrado

⁶ Merton critica os três postulados predominantes na análise funcional de Malinovski: os postulados da unidade funcional da sociedade, do funcionalismo universal e da indispensabilidade.

⁷ Merton (1968, p. 60) opõe as teorias de médio alcance aos “esforços sistemáticos e abrangentes para desenvolver uma teoria unificada que explique todas as uniformidades observadas no comportamento social, na organização social e na mudança social”.

⁸ Também é irônico observar que as duas tradições metodológicas iniciais trocaram sua afiliação institucional: o experimento floresceu em Chicago, enquanto metodologias de pesquisa inovadoras também foram desenvolvidas na Universidade de Columbia. Quanto à pesquisa qualitativa, desenvolvida anteriormente em Chicago, ela foi revivida pela avaliação somente mais tarde, com a “avaliação responsiva” de Stake e a “avaliação naturalista” de Guba e Lincoln; o último rótulo foi cunhado em oposição à ideia de que seria possível tratar os programas como experimentos feitos em um laboratório (Guba & Lincoln, 1987).

⁹ Campbell (1979, p. 84) recomendou que “os metodologistas de pesquisa de avaliação [...] [deveriam] se recusar a usar nossas habilidades em pesquisas ad hominem. Esses resultados são de generalização muito limitada. Nossas habilidades devem ser reservadas para a avaliação de políticas e programas que possam ser aplicados em mais de um ambiente e que qualquer administrador bem-intencionado com financiamento adequado possa adotar”.



por uma extensa literatura sobre experimentação social¹⁰. Os experimentalistas eram vistos como “especialistas metodológicos (que) se concentravam no problema de avaliar o impacto da mudança social” (Campbell, 1979, p. 67).

Os experimentalistas logo se depararam com a dificuldade de explicar se os resultados esperados eram de fato o efeito do programa, ou mesmo se os programas haviam produzido algum resultado de fato. A maioria das avaliações conduzidas com projetos experimentais mostrou quase nenhuma diferença positiva entre os grupos experimentais e de controle, uma situação desanimadora para as pessoas que defendiam a reforma social e que temiam que um resultado negativo pudesse ser utilizado pelos políticos para interromper os programas.

Havia duas maneiras de abordar o problema dos resultados ausentes. A primeira dizia respeito à substância. Como Rossi (1969) observou, os programas abordavam problemas cronicamente insolúveis, e talvez fosse simplesmente irrealista pensar que os experimentos poderiam funcionar como milagres. Isso colocou em questão a lógica dos programas existentes e levou a solicitações de melhores teorias para orientá-los: um caminho que Chen & Rossi (1983) seguiram mais tarde, com a proposta de avaliações “orientadas pela teoria”, mas que pode não ter parecido fácil na época.

A segunda abordagem era principalmente metodológica e atraiu mais atenção: se não fosse possível provar nada, o experimento não tinha sido feito corretamente. A maior parte da literatura sobre design experimental trata de tópicos como a seleção de grupos experimentais e de controle, validade estatística, medição do tratamento, e afins. Essa literatura reconhece as dificuldades de realizar experimentos em campo e, muitas vezes, chega a conclusões minimalistas: os experimentos podem ser realizados com sucesso quando há certas condições, como objetivos claros, uma necessidade real de inferência causal e programas pequenos¹¹.

Entretanto, os experimentalistas não podiam desconsiderar o impacto das teorias na metodologia. Suchman (1967) é frequentemente citado como tendo dito que um efeito nulo poderia ser atribuído tanto a uma falha na implementação do programa, quanto a uma falha na teoria do programa. Em uma revisão de 175 avaliações experimentais, Lipsey et al. (1985) agrupam os estudos em não teóricos (“tratamentos de caixa preta”), subteóricos (nos quais apenas a estratégia do programa ou os princípios do programa são mencionados) e teóricos (contendo “formulações específicas que vinculam elementos do programa aos resultados desejados”). A falta de uma teoria foi responsável pela má operacionalização das variáveis e, portanto, pela baixa qualidade da pesquisa¹².

O próprio Campbell sustentava que, se não fosse possível encontrar um vínculo causal entre a intervenção e o efeito, dever-se-ia procurar outras hipóteses plausíveis: se não fosse alcançado, o efeito desejado poderia ser buscado por meio de outra intervenção; se fosse alcançado, poderia ter sido resultado de outras causas que não a intervenção, como na redução das taxas de mortalidade após a introdução de limites de velocidade nas rodovias de Connecticut, analisada em Campbell (1968). Em vez de “nada funciona”, pode-se dizer que “algo mais funciona”. Campbell admitiu que, a menos que o promotor de um programa fosse também o “administrador de experimentos”, ele poderia não estar interessado em outro programa que

¹⁰Veja, por exemplo, Riecken & Boruch (1974), Bennet & Lumsdaine (1975).

¹¹Para um relato interessante dessa alegação minimalista, consulte a “Introdução” de Manski & Garfinkel (1992).

¹²Em outros casos, observou-se que o trabalho exigido dos analistas para estimar a distribuição de custos e benefícios de programas alternativos teria exigido um “modelo comportamental da população que possibilitasse simular os efeitos de políticas alternativas” (Rivlin, 1971, p. 32) para as quais as ferramentas existentes eram inadequadas. Com relação aos programas de manutenção de renda, Rivlin observou que seria necessário: primeira etapa, reproduzir as características de uma amostra de famílias de baixa renda; segunda etapa, basear-se em pesquisas e experimentos para obter estimativas dessas respostas comportamentais; terceira etapa, analisar as compensações entre a manutenção da renda e estratégias alternativas de investimento contra a pobreza. Em 1969, apenas a primeira etapa havia sido executada, e não era possível fazer qualquer “provisão para mudanças de comportamento que pudessem resultar de mudanças nos programas de manutenção de renda” (Rivlin, 1971, p. 34).



talvez pudesse ser implementado por adversários políticos no próximo mandato¹³. Naquela época, no entanto, pouquíssimos - mesmo entre seus colegas experimentalistas - seguiram Campbell na mensagem "visionária" contida em ver "reformas como experimentos".

A Escola de Columbia e o Bureau of Applied Social Research

A tradição sociológica que se desenvolveu em torno do Bureau of Applied Social Research desempenhou um papel menor na avaliação dos programas sociais da guerra contra a pobreza do que se poderia esperar, considerando seu histórico na década anterior. Os principais debates sobre avaliação, como aqueles sobre atribuição de causalidade ou sobre valoração, não foram iniciados pelo pessoal do Bureau. Além disso, a estrutura conceitual central do Bureau não estruturou o discurso da avaliação. No entanto, muitos elementos teriam previsto uma presença mais forte.

Primeiro, seu trabalho foi identificado com a pesquisa social aplicada, que é o que a avaliação é. Sua predileção pela pesquisa social aplicada chegou a colocar em risco sua posição acadêmica. Havia uma distinção bem estabelecida entre, de um lado, a pesquisa básica envolvida na descoberta de novas teorias e metodologias e, de outro, a pesquisa aplicada que repetia dispositivos de rotina, realizada com o mero objetivo de obter dinheiro para apoiar à primeira¹⁴. Em vez disso, o Bureau realizava pesquisas aplicadas para comissionados privados ou públicos, visando buscar seus interesses teóricos e metodológicos na "sociedade em movimento": a maioria de suas principais descobertas teóricas tinha surgido de estudos encomendados sobre mudanças sociais planejadas, como campanhas de persuasão em massa ou programas de educação e saúde.

Em segundo lugar, eles desenvolveram técnicas de pesquisa especialmente adaptadas para avaliação, alternativas aos projetos (quase) experimentais. Em particular, em "Principles of Evaluation", um capítulo do estudo pioneiro de avaliação "Encampment for Citizenship", de Hyman et al. (1962), sustentou-se que a avaliação está interessada em resultados inesperados (a "serendipidade" de Merton) tanto quanto nos esperados; que, como os programas eram administrados para pessoas que viviam no mundo real, não era possível observar grandes mudanças nos grupos experimentais e, portanto, era aconselhável realizar análises comparativas e longitudinais para detectar pequenas mudanças em toda a população visada pelo programa; que mecanismos como a influência pessoal deveriam ser usados como variáveis intervenientes e as variáveis de contexto deveriam ser trabalhadas; que os métodos de pesquisa qualitativa deveriam ser usados lado a lado com os quantitativos.

Terceiro, seu interesse teórico em compreender as mudanças no comportamento social estava de acordo com o tipo de problemas que a guerra contra a pobreza estava abordando. De fato, Lazarsfeld (1967) disse que o Bureau estava bem posicionado para realizar a avaliação porque combinava um conhecimento especializado em técnicas de pesquisa com uma consciência das variáveis sociológicas, como estratificação social, relações sociais primárias, grupos de referência e pressões cruzadas. No entanto, era mais fácil falar do que fazer isso. De fato, podemos sonhar com o que um forte endosso dessa declaração poderia ter produzido. Coleman (1990), que compartilhava dessa opinião, cita apenas alguns estudos que utilizaram essas ferramentas sociológicas e conseguiram superar o "erro da simples agregação" que assolava a pesquisa social aplicada comum¹⁵. Um estudo de Coleman et al. (1966) sobre o uso de um novo medicamento na prática médica utilizou a teoria de Katz e Lazarsfeld sobre os fluxos de comunicação em duas etapas; outro estudo de Schultze (1977)

¹³Outras correntes de pensamento abordaram o mesmo problema. Os economistas de escolha pública afirmam que há poucos incentivos para que os burocratas se envolvam em experimentos, pois isso prejudicaria seus esforços de maximização do orçamento. Por outro lado, considerando a oposição entre o "homem administrativo", ávido por estabilidade, e o "homem avaliador", ávido por mudanças, Aaron Wildavsky (1972) concebeu uma "organização autoavaliadora", em que a avaliação deveria ter entrado no jogo de redefinição de metas e de mostrar em que uma organização era boa, em vez de apenas observar que as metas não haviam sido cumpridas.

¹⁴Para conhecer a posição do Bureau sobre isso, consulte Merton (1949) e Lazarsfeld (1967).

¹⁵Ironicamente, pode-se dizer que o próprio Coleman não foi capaz de seguir esse impulso porque se voltou para o lado oposto: "grande teorização" (Horowitz, 1993, p. 109).



sobre incentivos enfatizou que “tanto a participação no programa [...] quanto a resposta ao programa dependem da estrutura social existente entre os receptores” (Coleman, 1990, p. 646).

Por que, então, a tradição do Bureau não conseguiu desempenhar o papel que poderia ter sido esperado quando a avaliação de grandes programas sociais estava em pauta?¹⁶ Por que, na presença de resultados inesperados e de contextos sociais variados, eles não foram capazes de propor uma abordagem de avaliação baseada em teorias de “médio alcance”? Para responder a essa pergunta, vamos primeiro nos voltar para a história do Bureau. Enquanto Merton e Lazarsfeld colaboraram, até meados da década de 1950, o diálogo inovador entre teoria e métodos de pesquisa exerceu um efeito positivo em tudo o que eles fizeram, e a avaliação do Encampment pôde se beneficiar dessa parceria frutífera. Entretanto, uma década depois, a chama não estava mais acesa e uma atitude tradicional havia ressurgido, reforçando o papel central da metodologia na definição da disciplina sociológica e de seus campos de aplicação. A guerra contra a pobreza estava em seu ápice quando pessoas do Bureau, como Hyman & Wright (1967, p. 743), cunharam a frase: “A avaliação é o acompanhamento metodológico da ação social planejada”. Em seu artigo programático sobre “Avaliação de programas de ação social”, foi dada atenção especial à conceituação das variáveis independentes e dependentes que constituíam o programa, concentrando-se no “manipulável e testável” (Hyman & Wright, 1967, p. 756): um plano de pesquisa totalmente preocupado com variáveis e desprovido de mecanismos¹⁷. De fato, o fervor otimista da época visava os impactos diretos e mensuráveis das intervenções (tratamentos, variável independente) e desconfiava das influências indiretas e dos mecanismos ocultos.

Além disso, foi um acadêmico treinado no Bureau, Suchman (1967), que reconheceu que o (quase) experimento havia vencido a batalha do método e se juntou ao campo dos experimentalistas. Se o método era o fator crucial, então, segundo o argumento, dever-se-ia seguir o “melhor” método, como o projeto experimental costumava ser considerado. Dessa forma, a metodologia de avaliação começou a ser julgada apenas por seus próprios méritos e não com base em sua adequação ao programa a ser avaliado; e os métodos foram ordenados hierarquicamente, desde os experimentos no topo até a observação na base, passando por vários tons de misturas qualitativas/quantitativas.

O que foi dito até agora explica a falta de um impacto direto sobre o progresso da avaliação pelo Bureau como um centro de pesquisa com uma tradição intelectual específica. Uma história diferente seria a de eminentes avaliadores treinados em Columbia que, embora não se identificassem com essa tradição, definitivamente mantiveram sua marca. Peter Rossi, coautor de *Evaluation: A Systematic Approach*¹⁸ (Rossi et al., 1979), foi aluno de pós-graduação em Columbia na década de 1940, mas suas primeiras experiências de avaliação ocorreram em Chicago, no National Opinion Research Center, onde foi influenciado principalmente pela metodologia (quase) experimental de Campbell (Rossi, 2004). Mais tarde, ele admitiu a dificuldade de realizar experimentos, falou sobre projetos de controle reflexivo e outras alternativas metodológicas para avaliar o impacto, e também se tornou um defensor da avaliação orientada pela teoria.

Carol Weiss escreveu sua pesquisa de avaliação (1972) na Columbia, mas foi exposta principalmente a influências da análise de políticas¹⁹. Weiss trabalhou com mais continuidade do que qualquer outro autor em temas que estavam no centro das atenções do Bureau, como os usos da pesquisa²⁰, o papel do conhecimento, as consequências não intencionais e os mecanismos. Em seu primeiro livro, no qual ela concentrou sua atenção no projeto de avaliação, a necessidade de teoria já estava presente: “um dos elementos mais importantes

¹⁶Uma exceção notável é o estabelecimento de um “prêmio Paul Lazarsfeld para um ensaio sobre teoria da avaliação” concedido pela American Evaluation Association.

¹⁷Para essa distinção, consulte Pawson (1989).

¹⁸A primeira edição de *Evaluation: A Systematic Approach* (Avaliação: Uma Abordagem Sistemática) teve como coautores H.E. Freeman e S. Wright, a segunda edição até a quinta edição somente H.E. Freeman. Na sexta e sétima edições, M. Lipsey entrou como coautor.

¹⁹Weiss (2004, p. 162) observa que os cientistas sociais que mais influenciaram seu trabalho foram Jim March (um teórico das organizações), Ed Lindblom (um economista/cientista político) e Lee Cronbach, alguém “mais próximo de casa para os avaliadores”.

²⁰Algo em que Merton trabalhava desde 1949, como Weiss citou com frequência.



na produção de uma avaliação útil é situar o estudo em uma perspectiva teórica. [...] Tem de haver alguma razão - alguma justificativa teórica - para esperar que um programa seja bem-sucedido" (Weiss, 1972, p. 84).

Evidência e complexidade

Trinta anos depois, as coisas mudaram radicalmente. Os programas sociais abrangeram quase todos os aspectos possíveis da vida social, e uma grande quantidade de conhecimento foi coletada sobre eles por meio de avaliações e outros tipos de pesquisa (Rist & Stame, 2006). No entanto, o otimismo deu lugar ao pessimismo; problemas sociais novos e antigos ainda precisam ser resolvidos, e há uma necessidade urgente de documentação - especialmente por parte de governos céticos - sobre o que funciona e por quê. Esse desafio foi aceito por acadêmicos e profissionais ligados de alguma forma às duas abordagens examinadas até agora. Eles forneceram novas percepções, mostrando continuidades com essas abordagens, mas também se afastando delas.

O movimento em prol de políticas baseadas em evidências faz um balanço da multiplicidade de intervenções que lidaram separadamente com um determinado problema e pretende avaliar quais funcionam e quais não funcionam - com a implicação de que os tomadores de decisão devem se concentrar nas primeiras e desistir das segundas. Paralelamente ao que aconteceu no campo da saúde com o estudo dos efeitos dos tratamentos médicos (a Colaboração Cochrane), novos grupos, como a Colaboração Campbell e outros centros de pesquisa especializados que se concentraram na transferência de conhecimento a partir de experimentos, começaram a produzir revisões sistemáticas de estudos anteriores sobre os efeitos de diferentes intervenções (incentivos e serviços) no campo social. Em seu esforço para dar sentido à enxurrada de material, eles estabeleceram regras para selecionar os estudos que haviam sido conduzidos com "metodologias sólidas": há um entendimento claro de que a evidência se baseia na metodologia e que há uma hierarquia de métodos, com os Ensaios de Controle Randomizados (RCT) no topo (Campbell Collaboration, 2001).

Isso, por sua vez, provocou um debate intenso²¹ sobre revisões, meta-análises e sínteses (Perrin, 2005) e deu um novo impulso ao diálogo entre teoria e método. Nas palavras de Mulgan (2003, p. 4), "as evidências não existem de forma abstrata e flutuante. Ela existe em relação a teorias e conceitos que fornecem o prisma pelo qual o mundo é visto. Essas teorias não são alternativas aos fatos e evidências concretas, elas são a única maneira de dar sentido a eles". No entanto, a Colaboração Campbell considera o lugar da teoria como algo marginal. Perguntas como "as teorias preveem como as principais variáveis envolvidas na revisão se relacionarão entre si?" ou "diferentes teorias ou filosofias de tratamento geram práticas conflitantes?" pertencem à "declaração de antecedentes" de uma revisão (Campbell Collaboration, 2001, p. 3) e não desempenham um papel importante na elaboração da revisão. De acordo com Oakley (1999, p. 5), embora "algumas intervenções, especialmente no campo social, sejam informadas por teorias sobre processos de intervenção e/ou de mudança de comportamento [...] se a teoria faz ou não diferença para a eficácia da intervenção é uma questão, em grande parte, sem resposta".

Uma resposta positiva a essa pergunta foi oferecida por uma perspectiva diferente sobre a unidade de análise das revisões, conforme proposto pelas sínteses realistas (Pawson, 2002). Há enormes dificuldades em fazer generalizações a partir de iniciativas existentes que se mostraram boas, mas que dificilmente poderiam ser reproduzidas na mesma "assembleia". Em vez disso, o que pode ser comparado são os mecanismos pelos quais os sujeitos reagiram ao estratagema de intervenção: as sínteses devem "reunir uma vasta experiência das opções e possibilidades e descobrir que tipo de coisas funcionam para que tipo de sujeitos e em que tipo de situações" (2002, p. 349). Esse conhecimento cumulativo, obtido por qualquer método possível, inclusive entrevistas realistas, resulta em uma teoria de médio alcance por meio de um processo de abstração que utiliza o conceito realista de causalidade generativa em oposição

²¹A revista Evidence and Policy oferece uma boa base para esse debate.



ao conceito experimentalista de causalidade sequencial²². Embora a formação sociológica de Pawson seja excêntrica no contexto da tradição dos EUA, sua contribuição é relevante aqui porque tem uma dívida com as teorias de médio alcance de Merton (consulte Pawson, 2010). No exemplo de uma variedade de intervenções que usam a estratégia de “nomear e envergonhar” (Pawson, 2006), a teoria de médio alcance dos grupos de referência foi usada para explicar as reações positivas, indiferentes ou hostis das pessoas inseridas em diferentes contextos com base em sua atitude em relação ao grupo que foi nomeado e envergonhado.

Uma segunda corrente de pensamento avaliativo concentrou-se na própria combinação de intervenções que podem ocorrer quando: a) existe uma gama de programas em um único local; b) os programas implementados em determinados locais são o resultado de coalizões verticais (gestão intergovernamental, subsidiariedade); c) os programas que visam a um único problema, como pobreza ou desemprego, têm muitos componentes (sociais, econômicos, ambientais, etc.); d) os programas reúnem forças individuais, coletivas e globais ao mesmo tempo (parcerias). A complexidade é o conceito abrangente para todas essas situações: é amplamente reconhecido que ela desafia a possibilidade de atribuir causalidade por meio de controles randomizados e análises contrafactuais, que são necessárias abordagens de avaliação alternativas adequadas.

Aqui está mais uma oportunidade para que as potencialidades ocultas da tradição da Escola de Columbia venham à tona. A complexidade não é um fenômeno novo. Ela sempre esteve associada à necessidade de abandonar as generalizações monocausais. Ouça o que Albert Hirschman (1995, p. 116) diz sobre sua descoberta de estratégias alternativas de desenvolvimento que desafiavam a lógica dos programas convencionais: “Foi uma generalização ou teoria da ‘faixa intermediária’, como Robert Merton a chamou. Ao lidar com os múltiplos e complexos problemas do desenvolvimento, aprendemos que devemos criar generalizações em todos os tipos de faixas e ser surdos, como Ulisses, ao canto sedutor do paradigma único”.

Entretanto, a visualização da complexidade tornou-se o ponto de virada das novas elaborações. A abordagem baseada em teoria de Carol Weiss originou-se da necessidade de compreender a multiplicidade de mecanismos que poderiam explicar como funcionavam os programas implementados em ambientes complexos, como o Comprehensive Community Initiatives (objeto de uma pesquisa dirigida por Weiss). Weiss (1995) enumera as diferentes teorias que podem ser levadas em consideração: não apenas as teorias que podem ter sido escritas no programa (as consideradas por Lipsey et al., 1985), mas também aquelas que podem ter orientado os programadores ou que podem ser elaboradas pelos avaliadores, sem mencionar as teorias tácitas das partes interessadas. O interesse especial de Weiss é entender os mecanismos que fazem as pessoas reagirem aos programas. Para isso, ela distingue entre teorias de implementação e teorias de programas: as primeiras explicam o que acontece em um programa, o processo à medida que ele se desenvolve passo a passo. A segunda explica o que acontece entre a implementação e o desencadeamento dos efeitos, ou seja, os mecanismos de trabalho responsáveis pelo sucesso de um programa. Esses mecanismos, de acordo com Weiss (1997), podem ser de natureza cognitiva, afetiva, social ou organizacional, e ela convida os avaliadores a procurar teorias nessas disciplinas relacionadas e a encontrar maneiras de analisar e medir como elas podem ser detectadas; os métodos se tornam relevantes na medida em que apoiam o processo de elaboração da teoria.

Discussão

Na era dos grandes programas sociais da Great Society, incluindo a guerra contra a pobreza, a tradição sociológica dos EUA ofereceu à avaliação uma metodologia para testar as hipóteses ocultas de causalidade linear, segundo as quais os programas sociais deveriam funcionar em nível geral. As duas principais correntes sociológicas da época - os experimentalistas e a escola de Columbia - concentraram-se na metodologia para testar o impacto da variável independente (o programa) sobre a variável dependente (os beneficiários, o alvo) e para

²²O que muda quando um programa é implementado não é o status de uma característica (variável) do beneficiário que recebe passivamente o tratamento, como na causalidade sequencial, mas o ator que está inserido em uma situação e reage ao programa acionando um mecanismo (causalidade geradora).



medir a extensão de seu sucesso. Em ambos os casos, os especialistas em ciências sociais não trouxeram seus conhecimentos teóricos para o planejador “racional”; em vez disso, eles se concentraram na função auxiliar do metodologista, cuja tarefa era testar se intervenções únicas sempre poderiam produzir os efeitos pretendidos.

Essa orientação mostrou seus limites desde o início: o problema do efeito nulo para os experimentalistas e o dos mecanismos mediadores para a escola de Columbia não podiam ser resolvidos apenas com métodos sofisticados. Foi somente quando a questão da complexidade surgiu como crucial que ambas as tradições conseguiram encontrar um avanço em seu próprio legado intelectual.

No campo da experimentação social, o argumento de Campbell sobre hipóteses plausíveis rivais oferecia um porto seguro. Além de levantar o problema da validade externa, o fato de que um determinado resultado poderia ser atribuído a outras causas que não o tratamento - sustentava Campbell - não era tão ruim, mas sugeria mudar o foco da intervenção para explicar o que havia sido descoberto como mais eficaz. Embora essa atitude ousada seja contrária à disposição atual dos defensores da Política Baseada em Evidências²³, vale a pena observar que um grande defensor dos Ensaios de Controle Randomizados, como Cook, expressou a opinião de que a avaliação deve ser “muito mais do que métodos de ciências sociais aplicados ao estudo de intervenções sociais” e comparou o campo da avaliação a um banquinho com três pernas: método, teoria (de avaliação) e compêndios de descobertas substantivas (Cook, 1997).

A tradição da Escola de Columbia, por outro lado, tem enfrentado a dificuldade de conciliar suas próprias almas conflitantes: a centralidade do método e a necessidade de descobrir os mecanismos em ação na “sociedade em movimento”. Na época da Guerra contra a Pobreza, quando os principais projetos de avaliação assumiam como hipóteses as teorias lineares por trás dos programas, eles não eram capazes de desafiá-las com base no que já se sabia sobre o efeito mediador da inserção social e da influência pessoal no comportamento social.

No entanto, mesmo que essa oportunidade não tenha sido aproveitada na época, e mesmo que não tenha sido reconhecida como um produto da escola de Columbia quando foi adotada por Carol Weiss (1972), as sementes estavam lá e deram frutos mais tarde, quando o reconhecimento da complexidade das situações em que os programas foram implementados exigiu uma compreensão da mudança social como um resultado combinado de efeitos intencionais e não intencionais da ação social proposital.

Atualizações

Mais de quinze anos depois, as abordagens orientadas por métodos e as abordagens baseadas em teorias continuam se confrontando, sob o nome de “política baseada em evidências” (EBP, abreviação do inglês; mesmo nome, perspectiva alterada) e “avaliação da teoria da mudança”.

Abordagens baseadas em evidências

O movimento em prol da política baseada em evidências está em andamento, mas seus defensores passaram das grandes sínteses para a pequena dimensão, esquecendo-se da visão ampla de Campbell e da cautela de Cook. A tocha da “metodologia padrão-ouro” está agora nas mãos de grupos de pesquisa como o 3IE (Iniciativa Internacional para Avaliação de Impacto), que agitou a bandeira da “avaliação de impacto rigorosa”, em que o impacto é definido como: “a diferença no indicador de interesse (Y) com a intervenção (Y1) e sem a intervenção (Y0)” (White, 2010). Em p. 154 outras palavras, uma definição limitada de impacto²⁴, de volta à definição de causalidade como o efeito líquido que pode ser atribuído a um programa específico (causa).

²³Não é de surpreender que Pawson (2004) tenha sugerido que Campbell não faria parte da Campbell Collaboration.

²⁴Para uma definição limitada versus ampla de impacto, consulte Stern et al. (2012, p. 20). A definição ampla é a definição clássica da OCDE: “efeitos positivos e negativos, primários e secundários, de longo prazo, produzidos por uma intervenção de desenvolvimento, direta ou indiretamente, intencional ou não intencional”, e não se baseia na metodologia da avaliação, como na definição limitada, mas no conteúdo do programa, para o qual a abordagem de avaliação deve ser apropriada (princípio da capacidade de resposta situacional).



Uma ideia relacionada é que é possível afirmar se uma política obteve os resultados esperados se ela foi projetada de forma a ser avaliada (dados de linha de base, criação de grupos de controle, etc.). O Poverty Action Lab, dirigido por Abhijit Banerjee e Esther Duflo (2011), uma contribuição pela qual eles ganharam o prêmio Nobel de Economia em 2019, levou esse movimento a consequências extremas ao sustentar que os programas devem ser concebidos de forma a serem avaliados por RCTs, limitando assim enormemente o tipo de programas que têm alguma chance de serem promovidos: fornecimento de mosquiteiros ou estabelecimento do melhor tamanho da sala de aula são os projetos emblemáticos. Essa maneira específica de abordar o desafio da complexidade, com base em um critério metodológico, corresponde a um descontentamento persistente com os resultados dos programas de desenvolvimento social. Como Woolcock (2023) sugeriu, já que, de acordo com esses autores, os grandes programas fracassam, é melhor concentrar todos os esforços econômicos em pequenos projetos.

A inspiração metodológica desse argumento provocou dois tipos diferentes de reações disciplinares. Os filósofos contestaram o núcleo da teoria da causalidade que sustenta essa abordagem: Cartwright & Munro (2010) observaram que os experimentos podem responder à pergunta “funcionou aqui?”, mas não “funcionará para nós (em outro lugar)?” (a questão da validade interna/externa que Campbell já havia abordado).

Alguns economistas, ao aprová-lo, sentiram que finalmente tinham um lugar na mesa de avaliação. No entanto, outros economistas, e não menos importantes (por exemplo, Deaton, que recebeu o Prêmio Nobel de Economia em 2015), levantaram a voz contra a usabilidade do RCT: “Descobrir o que funciona, mesmo que isso seja possível, é, por si só, uma base profundamente inadequada para a política” (Deaton, 2020, p. 2).

Esse tipo de crítica não deixou o campo da política baseada em evidências indiferente. Mesmo dentro da 3IE, Howard White admite²⁵: “A força dos Ensaios de Controle Randomizados (RCTs) é sua capacidade de estabelecer relações causais entre a intervenção e o resultado. Mas precisamos de uma análise factual do que aconteceu, para complementar a análise contrafactual da causalidade”. E, para esse fim, “nos preocupamos com as Teorias da Mudança”.

Teorias da mudança

A abordagem Avaliações Baseadas em Teorias (ABT) ganhou enorme força sob o pretexto da Teoria da Mudança (TdM) como uma ferramenta necessária para programas e avaliações. Aqui a tradição de Merton de consequências não intencionais está conhecendo um momento de grande popularidade.

O conceito de “Teoria da Mudança” foi proposto por Connell & Kubisch (1995) como uma contribuição para a pesquisa sobre Iniciativas Comunitárias Abrangentes (CCI, veja acima), mas mais tarde foi destacado como uma abordagem de avaliação própria. As TdM são “estudos sistêmicos e cumulativos das ligações entre atividades, resultados e contextos da iniciativa” (Connell & Kubisch, 1995, pp. 2-3). Elas têm “várias vertentes - econômica, política, social - que operam em vários níveis - comunitário, institucional, rede pessoal, familiar e individual”. Várias teorias de mudança podem estar operando simultaneamente em um único (programa) e várias partes interessadas podem estar trabalhando com teorias de mudança diferentes e possivelmente até concorrentes. Pode haver TdM explícitas ou implícitas, mas a avaliação precisa elaborar uma Teoria da Mudança funcional (“plausível, factível e testável”). As TdM visam entender como um programa funciona, para quem e por quê²⁶. Assim, investigam as várias maneiras que um programa pode ter seguido para produzir resultados esperados ou não esperados.

A grande influência da abordagem da TdM fez com que os avaliadores respondessem a uma série de perguntas diferentes: (1) Quem deve elaborar a TdM? O programa em si, o comissário, o avaliador, várias partes interessadas em conjunto? O que acontece quando diferentes partes interessadas elaboram TdM diferentes? (2) Para que serve uma TdM?

²⁵ Blog do 3ie (2023).

²⁶ As ToCs podem ser simplesmente representadas por gráficos em que as caixas (intervenção, efeito) são ligadas por setas (ligação causal) ou, de forma mais ampla, expressas por meio de narrativas do que está sendo observado ou do que se espera que aconteça.



É uma hipótese a ser verificada? Ou é um insumo para iniciar um processo de reflexão sobre o que está acontecendo, o que é esperado e/ou inesperado?

Há duas maneiras principais de abordar essas questões. Uma primeira tendência, muito popular, preocupa-se principalmente com o que é uma boa “teoria” ou, nas palavras de Mayne (2017), com o que é uma TdM “robusta”. Aqui aparece o impulso metodológico, com duas ideias principais: 1) que uma teoria robusta se baseia em “boas suposições” e 2) que, no mundo complexo atual, não é possível atribuir um efeito a uma única causa (o programa), mas que uma boa teoria é uma teoria de contribuição, em que mais causas contribuem para o efeito, e o programa é visto como uma possível causa contribuinte. Ao contrário dos RCT, na análise de contribuição, a questão da causalidade trata de encontrar a causa (contributiva) dos efeitos “observados”, não os efeitos “desejados” de uma determinada causa (atributiva). Para isso, está aberta a diferentes metodologias e abordagens, que privilegiam a análise do que está acontecendo em detrimento da predisposição de indicadores do que deveria acontecer: é o caso da “Análise Comparativa Qualitativa”, “rastreamento de processos”, “mapeamento de resultados” (Befani & Mayne, 2014). Nesse uso, uma TdM é principalmente de responsabilidade do avaliador, embora ele possa consultar outras partes interessadas.

Uma segunda tendência é aquela que se concentra nas “mudanças” que estão ocorrendo durante o processo de implementação. Aqui, a ideia é que uma TdM deve ser elaborada em conjunto pelas várias partes interessadas, e que também pode haver mais de uma TdM por vez. Valters (2014, pp. 9-13) argumenta da seguinte forma: “As teorias de mudança raramente se desenvolvem como previsto. Elas precisam ser adaptadas e retrabalhadas à medida que surgem novas informações (...) Posso ter três versões de TdM no mesmo dia”. Os avaliadores estão “tentando esclarecer constantemente e desafiar a si a dizer ‘é isso que entendemos que está acontecendo?’”, desempenhando assim uma “função desafiadora em que a TdM poderia ser regularmente avaliada de forma crítica tanto pelos membros da equipe quanto por vozes externas”: no final, a TdM tem o “suposto objetivo de desafiar as suposições dos atores do desenvolvimento”.

Embora ambas as tendências tenham em comum a ideia de complexidade, enquanto a primeira está empenhada em entender a combinação de muitas consequências “esperadas” (graças a boas suposições) das várias causas contribuintes, a segunda deixa mais espaço para a ideia de consequências inesperadas, sejam elas positivas ou negativas. E também dá mais espaço para o protagonismo dos vários atores. Dessa forma, ela é mais bem servida por abordagens de avaliação da persuasão do pensamento positivo, como “mudança mais significativa”, “desvio positivo”, “investigação apreciativa” (Stame, 2014).

Fonte de financiamento

Não há

Conflito de interesse

Uma primeira edição deste texto foi publicada em 2010, na décima sexta edição de *Mind the Gap*, ou “cuidado com o vão” (tradução livre), coleção voltada a elevar a atenção dos agentes com a implementação de iniciativas e políticas públicas em cenários reais. A referência bibliográfica daquela edição é Stame (2010). A pedido do editor-chefe da *Revista Brasileira de Avaliação*, o ensaio foi rerepresentado agora, preservando seu núcleo original, mas ganhando textos de apresentação e conclusão. O texto foi também traduzido com vistas a melhor alcançar a comunidade avaliativa de língua portuguesa. Todo o processo foi de comum acordo com os editores da primeira versão do ensaio.

Agradecimentos

O editor-chefe agradece a Tayná Mendes, pelo trabalho de tradução. Tayná é Socióloga. Mestranda em Sociologia pelo PPGSA/UFRJ. Pesquisadora do NIED/IFCS. Atua na área de pesquisa, avaliação e monitoramento em educação e desigualdade.



Referências

- Banerjee, Abhijit, & Duflo, Esther. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. USA: Public Affairs.
- Befani, Barbara, & Mayne, John. (2014). Process tracing and contribution analysis: a combined approach to generative causal inference for impact evaluation. *IDS Bulletin*, 45(6), 17-36. <http://dx.doi.org/10.1111/1759-5436.12110>
- Bennett, Charles A., & Lumsdaine, Arthur A. (1975). *Evaluation and experiment*. Cambridge: Academic Press.
- Campbell Collaboration. (2001). *Campbell systematic reviews. Guidelines for the preparation of review protocols*. Recuperado em 18 de agosto de 2006.
- Campbell, Donald T. (1968). Reforms as experiments. *The American Psychologist*, 24(4), 409-429. <http://dx.doi.org/10.1037/h0027982>
- Campbell, Donald T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67-90. [http://dx.doi.org/10.1016/0149-7189\(79\)90048-X](http://dx.doi.org/10.1016/0149-7189(79)90048-X)
- Caro, Francis G. (Ed.). (1971). *Readings in evaluation research*. New York: Russel Sage Foundation.
- Cartwright, Nancy, & Munro, Eileen. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, 16(2), 260-266. PMID:20367845. <http://dx.doi.org/10.1111/j.1365-2753.2010.01382.x>
- Chen, Huey T., & Rossi, Peter H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, 7(3), 283-302. <http://dx.doi.org/10.1177/0193841X8300700301>
- Coleman, James S. (1990). *Foundations of social theory*. Cambridge: Harvard University Press.
- Coleman, James S., Katz, Elihu, & Menzel, Herbert. (1966). *Medical innovation*. Indianápolis: Bobbs-Merrill.
- Connell, James P., & Kubisch, Anne C. (1995). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (pp. 1-36). Washington: Aspen Institute.
- Cook, Thomas D. (1997). Lessons learned in evaluation over the past 25 years. In E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st Century* (pp. 30-52). Thousand Oaks: Sage Publications. <http://dx.doi.org/10.4135/9781483348896.n2>.
- Deaton, Angus. (2020). *Randomization in the Tropics revisited: A theme and eleven variations*. Cambridge: National Bureau of Economic Research. <http://dx.doi.org/10.3386/w27600>
- Guba, E. and Lincoln, Y. (1987) "The Countenances of Fourth Generation Evaluation." in: D. Palumbo (ed.) *The politics of program evaluation*, Beverly Hills, Sage Publications. Pp.292-234
- Hirschman, Albert O. (1995). *A propensity for self-subversion*. Cambridge: Harvard University Press.
- Horowitz, Irving L. (1993). *The decomposition of sociology*. Oxônia: Oxford University Press. <http://dx.doi.org/10.1093/oso/9780195073164.001.0001>.
- Hyman, Herbert, & Wright, Charles R. (1967). Evaluating social action programs. In P. F. Lazarsfeld (Ed.), *The uses of sociology* (pp. 741-782). New York: Basic Books.
- Hyman, Herbert, Wright, Charles R., & Hopkins, Thomas H. (1962). *Applications of methods of evaluation: four studies of the encampment for citizenship*. Berkeley: University of California Press. <http://dx.doi.org/10.1525/9780520321908>
- International Initiative for Impact Evaluation – 3ie. Recuperado em 18 de agosto de 2006, de <https://www.3ieimpact.org/taxonomy/term/771>
- Lazarsfeld, Paul F. (1967). Introduction. In P. F. Lazarsfeld (Ed.), *The uses of sociology*. New York: Basic Books.
- Leeuw, Frans L. (2003). Reconstructing program theories: methods available and problems to be solved. *The American Journal of Evaluation*, 24(1), 5-20. <http://dx.doi.org/10.1177/109821400302400102>
- Lipsey, Mark W., Crosse, Scott, Dunkle, Julie, Pollard, John, & Stobart, Gordon. (1985). Evaluation: The state of the art and the sorry state of the science. In D. S. Cordray (Ed.), *Utilizing prior research in evaluation planning*, (New Directions for Program Evaluation, Vol. 27, pp. 7-28). Hoboken: Jossey-Bass.
- Manski, Charles F., & Garfinkel, Irwin. (1992). *Evaluating welfare and training programs*. Cambridge: Harvard University Press.
- Mark, Melvin, Henry, Gary H., & Julnes, George. (2000). *Evaluation. An integrated framework for understanding, guiding, and improving public and nonprofit policies and programs*. Hoboken: Jossey Bass.
- Martire, Francesco. (2006). *Come Nasce e Come Cresce una Scuola Sociologica*. Merton, Lazarsfeld e il Bureau. Acireale: Bonanno Editore.
- Mayne, John. (2017). Theory of change analysis: building robust theories of change. *The Canadian Journal of Program Evaluation*, 32(2), 155-173. <http://dx.doi.org/10.3138/cjpe.31122>
- Merton, Robert K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review*, 1(6), 894-904. <http://dx.doi.org/10.2307/2084615>
- Merton, Robert K. (1949). *Social theory and social structure*. USA: Free Press.



- Merton, Robert K. (1968). On sociological theories of the middle range. In R. K. Merton (Ed.), *Social theory and social structure* (3rd ed., pp. 39-72). USA: Free Press.
- Mulgan, Geoff. (2003). Government, knowledge and the business of policy-making. *Canberra Bulletin of Public Administration*, 108, 1-5.
- Oakley, Ann. (1999, July). An infrastructure for assessing social and educational interventions: same or different? In *Background Paper for Meeting at the School of Public Policy*. London: University of London.
- Oakley, Ann. (2000). *Experiments in Knowing*. USA: The New Press.
- OECD-DAC (2010). Quality Standards for Development Cooperation. DAC Guidelines and Reference Series. <http://www.oecd.org/dataoecd/55/0/44798177.pdf>
- Pawson, Ray. (1989). *A measure for measures*. London: Routledge.
- Pawson, Ray. (2002). Evidence-based policy: The promise of realist synthesis. *Evaluation*, 8(3), 340-358. <http://dx.doi.org/10.1177/135638902401462448>
- Pawson, Ray. (2004). Would Campbell be a member of the Campbell collaboration? *The Evaluator, Winter*, 13-15.
- Pawson, Ray. (2006). *Evidence-based policy: a realist perspective*. Thousand Oaks: Sage Publications. <http://dx.doi.org/10.4135/9781849209120>
- Pawson, R. (2010) "Middle Range Theory and Program Theory Evaluation: from Provenance to Practice", in J. Vaessen and F.L. Leeuw (eds.) *Mind the Gap: perspectives on Policy Evaluation and the Social Sciences*, Transaction Publishers, New Brunswick, NJ. pp. 171-202
- Pawson, Ray, & Tilley, Nick. (1997). *Realistic evaluation*. Thousand Oaks: Sage Publications.
- Perrin, Burt (2005) "How evaluation can help make knowledge management real" in: Rist C.R & Stame N., eds., *From Studies to streams*, New Brunswick NJ: Transaction Publishers, pp.23-45
- Riecken, H.W., Boruch, R.F., Campbell D.T., Glennan T.K., Pratt, J., Rees A. & Williams W. (1974) *Social experimentation. A method for planning and evaluating social Interventions*, New York, Academic Press
- Rist, Ray C., & Stame, Nicoletta. (Eds.). (2006). *From studies to streams: Managing evaluative systems*. Piscataway: Transaction Publishers.
- Rivlin, Alice. (1971). *Systematic thinking for social action*. Washington: The Brookings Institution.
- Rossi, Peter H. (1969). Practice, method and theory in evaluating social-action programs. In J. L. Sundquist (Ed.), *On fighting poverty*. New York: Basic Books.
- Rossi, Peter H. (2004). My views of evaluation and their origins. In M. C. Alkin (Ed.), *Evaluation Roots — Tracing Theorists' Views and Influences* (pp. 122-131). Thousand Oaks: Sage Publications. <http://dx.doi.org/10.4135/9781412984157.n7>
- Rossi, Peter H., Freeman, Howard E., & Wright, Sonia R. (1979). *Evaluation: A systematic approach* (1st ed.). Thousand Oaks: Sage Publications.
- Schultze, Charles L. (1977). *The public use of private interest*. Washington: The Brookings Institution.
- Stame, Nicoletta. (2004). Theory based evaluations and types of complexity. *Evaluation*, 10(1), 58-76. <http://dx.doi.org/10.1177/1356389004043135>
- Stame, Nicoletta. (2010). US sociology and evaluation: Issues in the relationship between methodology and theory. In J. Vaessen & F. L. Leeuw (Eds.), *Minding the gap: Perspectives on policy* (Comparative Policy Evaluation, 16th ed., Vol. 3, pp. 29-44). USA: Transaction Publishers.
- Stame, Nicoletta. (2014). Positive thinking approaches to evaluation and program perspectives. *The Canadian Journal of Program Evaluation*, 29(2), 67-86. <http://dx.doi.org/10.3138/cjpe.29.2.67>
- Stern, Elliot, Stame, Nicoletta, Mayne, John, Forss, Kim, Davies, Rick, & Befani, Barbara. (2012). *Broadening the range of designs and methods for impact evaluation* (DFID Working Paper, No. 38, pp. 1-92). London: DFID. <http://dx.doi.org/10.22163/fteval.2012.100>
- Stern, Elliot. (2005). Introduction. In E. Stern (ed.), *Evaluation Research Methods* (Vol. 1, pp. XXI-XLIII). Thousand Oaks: Sage Publications. <http://dx.doi.org/10.4135/9781446261606>
- Stouffer, Samuel A. (1949). *The American soldier*. Princeton: Princeton University Press.
- Suchman, Edward A. (1967). *Evaluative research*. New York: Russel Sage Foundation.
- Vaessen, Jos, & Leeuw, Frans L. (Eds.). (2010). *Mind the gap: Perspectives on policy evaluation and the social sciences*. New Brunswick: Transaction Publishers.
- Valters, Craig. (2014). *Theories of change in international development: Communication, learning, or accountability?* London: JSRP, Asia Foundation.
- Weiss, Carol H. (1972). *Evaluation research*. Englewood Cliffs: Prentice Hall.
- Weiss, Carol H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives* (Vol. 1, pp. 65-92). Washington: The Aspen Institute.
- Weiss, Carol H. (1997). Theory-based evaluation: Past, present, and future. In D. J. Rog (Ed.), *Progress and future directions in evaluation, new directions for evaluation* (Vol. 76, pp. 41-55). San Francisco: Jossey-Bass. <http://dx.doi.org/10.1002/ev.1086>



Weiss, Carol H. (2004). Rooting for evaluation: A cliff notes version of my work. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influence* (pp. 153-168). Thousand Oaks: Sage Publications. <http://dx.doi.org/10.4135/9781412984157.n9>

White, Howard. (2010). A contribution to current debates in impact evaluation. *Evaluation*, 16(2), 153-164. <http://dx.doi.org/10.1177/1356389010361562>

Wildavsky, Aaron. (1972). The self-evaluating organization. *Public Administration Review*, 5, 509-520. <https://doi.org/10.2307/975158>

Woolcock, Michael. (2023). *International development: Navigating Humanity's greatest challenge*. Cambridge: Polity Press.