

Avaliação para o investimento social privado: metodologias

ORGANIZAÇÃO
FUNDAÇÃO ITAÚ SOCIAL
FUNDAÇÃO ROBERTO MARINHO
FUNDAÇÃO MARIA CECÍLIA SOUTO VIDIGAL
MOVE

Fundação **Santillana**



Avaliação para o investimento social privado: metodologias

II Seminário Internacional

“Avaliação para o Investimento Social Privado: Metodologias”

Rio de Janeiro, 3 de julho de 2013.

Organização

Fundação Itaú Social

Fundação Roberto Marinho

Fundação Maria Cecília Souto Vidigal

Parceria

Move Avaliação e Estratégia em Desenvolvimento Social

Apoio

Fundação Santillana

GIFE – Grupo de Institutos, Fundações e Empresas

Avaliação para o investimento social privado: metodologias

ORGANIZAÇÃO

FUNDAÇÃO ITAÚ SOCIAL
FUNDAÇÃO ROBERTO MARINHO
FUNDAÇÃO MARIA CECÍLIA SOUTO VIDIGAL
MOVE

Fundação **Santillana**



Avaliação para o investimento social privado: metodologias

© 2013 Fundação Santillana

Organização

Fundação Itaú Social (Antonio Bara Bresolin)

Fundação Roberto Marinho (Rosalina Soares)

Fundação Maria Cecília Souto Vidigal (Eduardo Marino)

Move Avaliação e Estratégia em Desenvolvimento Social (Rogério Silva)

Tradução

Babeth Bettencourt (artigo de Donna M. Mertens)

e Dolores Montero (artigo de Miguel Székely)

Revisão técnica das traduções

Rogério Silva

Produção editorial

Fundação Santillana

Diretoria de Relações Institucionais

Luciano Monteiro

Lucia Jurema Figueirôa

Edição

Ana Luisa Astiz

Projeto Gráfico

Paula Astiz

Editoreção Eletrônica e Gráficos

Laura Lotufo / Paula Astiz Design

Revisão

Cida Medeiros, Juliana Caldas e Katia Shimabukuro

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Avaliação para o investimento social privado : metodologias / organização
Fundação Itaú Social, Fundação Roberto Marinho, Fundação Maria Cecília Souto
Vidigal, Move . — São Paulo : Fundação Santillana, 2013.

Vários autores.
ISBN 978-85-63489-16-6

1. Ação social 2. Desenvolvimento social 3. Empresas - Aspectos sociais
4. Participação social 5. Pesquisa social - Metodologia 6. Projetos sociais
7. Seminário Internacional "Avaliação para o investimento social privado :
metodologias" (2013: Rio de Janeiro) I. Fundação Itaú Social. II. Fundação
Roberto Marinho. III. Fundação Maria Cecília Souto Vidigal. IV. Move.

13-11636

CDD-361.760981

Índices para catálogo sistemático:
1. Brasil : Avaliação para o investimento social privado :
Bem-estar social 361.760981

- 7 **Apresentação**
VALÉRIA VEIGA RICCOMINI, MÔNICA DIAS PINTO E EDUARDO QUEIROZ
- 11 **Avaliação como aprendizagem**
ANDRE DEGENSZAJN
- 17 **A escolha de métodos para fazer avaliação**
ROGÉRIO SILVA E DANIEL BRANDÃO
- 29 **Métodos mistos de avaliação:
aumentando a efetividade do investimento
social privado**
DONNA M. MERTENS
- 73 **Vantagens e complexidades das avaliações
de impacto**
MIGUEL SZÉKELY
- 93 **Avaliações válidas, relevantes e úteis**
THOMAZ KAUARK CHIANCA
- 117 **O uso da abordagem participativa
na avaliação do Projeto Maleta Infância**
FUNDAÇÃO ROBERTO MARINHO – CANAL FUTURA
- 125 **O uso de métodos mistos na avaliação
do Programa Primeiríssima Infância**
FUNDAÇÃO MARIA CECÍLIA SOUTO VIDIGAL
- 135 **Abordagens quantitativas na avaliação
de impacto do Programa Jovens Urbanos**
FUNDAÇÃO ITAÚ SOCIAL
- 143 **Anexo – Oportunidades para atualização
e publicação na área de avaliação**
FUNDAÇÃO ROBERTO MARINHO E INSTITUTO FONTE
- 159 **Agradecimentos**

Apresentação

VALÉRIA VEIGA RICCOMINI

Superintendente – Fundação Itaú Social.

MÔNICA DIAS PINTO

Gerente de desenvolvimento institucional – Fundação Roberto Marinho.

EDUARDO QUEIROZ

Diretor-presidente – Fundação Maria Cecília Souto Vidigal.

Este livro reúne um conjunto de contribuições oriundas do Seminário Internacional **Avaliação para o Investimento Social Privado: Metodologias**, realizado em julho de 2013 no Rio de Janeiro, pela Fundação Itaú Social, Fundação Roberto Marinho e Fundação Maria Cecília Souto Vidigal, em parceria com a Move e com apoio da Fundação Santillana e do GIFE - Grupo de Institutos, Fundações e Empresas.

A parceria entre essas instituições vem gerando subsídios que enriquecem o debate e as práticas de avaliação de projetos e programas sociais, agenda que ganha cada vez mais importância no País. Em 2012 tivemos a primeira edição do Seminário Internacional, intitulado **A Relevância da Avaliação para o Investimento Social Privado**, evento a partir do qual foi elaborada uma publicação com as reflexões produzidas no decorrer das apresentações e debates.

Na edição de 2013, o tema das metodologias ganhou destaque. Partimos do princípio de que as escolhas metodológicas são essenciais para que os processos avaliativos produzam informações relevantes e úteis a respeito dos resultados e impactos do investimento social privado. Ressaltamos, contudo, que a escolha de uma ou de outra abordagem metodológica depende dos objetivos e das questões que se pretende responder com a avaliação, e este é um enfoque que procuramos abordar nesta publicação.

Com o evento que realizamos, promovemos o diálogo entre diferentes pontos de vista metodológicos e entre perspectivas nacionais e internacionais, além de alcançar frutífera troca de experiências entre os diversos participantes. Dessa forma, aprofundou-se a compreensão sobre os processos avaliativos e de como os resultados de projetos sociais podem ser relevantes para o desenvolvimento social.

Nesta publicação, reunimos artigos e casos que retratam a diversidade de metodologias, suas potencialidades e limites, com o intuito de subsidiar escolhas diante de indagações como “Quando usar determinada metodologia?”, “Como equilibrar diferentes metodologias?”, “Qual metodologia é a mais efetiva para os objetivos que se pretende alcançar?”. Diante dessas e de outras perguntas pertinentes ao tema, o artigo de Andre Degenszajn, diretor executivo do GIFE, problematiza algumas características da cultura de avaliação no campo do investimento social privado. Na sequência, o artigo de Rogério Silva e Daniel Brandão, da Move, discorre sobre alguns dos principais fatores envolvidos nas escolhas metodológicas nos processos avaliativos, da ética à política, da técnica à viabilidade financeira.

Em seguida, o leitor encontrará os artigos produzidos pelos três palestrantes convidados para o seminário, que ganham o centro da publicação. Miguel Székely, doutor em Economia pela Universidade de Oxford, mestre em Políticas Públicas pelo Instituto Tecnológico Autónomo de México e diretor do Instituto para Inovação em Educação, situado em Monterrey, aborda a perspectiva econométrica. Donna M. Mertens, doutora em Psicologia da Educação pela Universidade de Kentucky, ex-presidente e atual conselheira da American Evaluation Association e professora da Universidade de Gallaudet, explica os métodos mistos.

Já Thomaz Kauark Chianca, doutor em Avaliação pela Western Michigan University, odontólogo da Universidade Federal do Rio de Janeiro (UFRJ) e avaliador com longa trajetória em consultorias no Brasil e no exterior, trata da produção de avaliações válidas, relevantes e úteis.

A fim de oferecer ao leitor um retrato de práticas avaliativas brasileiras, três estudos são apresentados na sequência que en-

cerra o livro. A avaliação do projeto Jovens Urbanos, da Fundação Itaú Social, retrata o uso da abordagem quantitativa para avaliar impacto; a avaliação do Projeto Maleta Infância, do Canal Futura, demonstra um processo de avaliação participativa; e o caso do Projeto Primeiríssima Infância, da Fundação Maria Cecília Souto Vidigal, mostra o uso de métodos mistos.

Para aqueles que estiverem interessados em aprofundar seu conhecimento no campo da avaliação, preparamos ainda um levantamento de **Oportunidades para atualização e publicação na área de avaliação**. Anexo ao livro, é um material que poderá inspirar leitores e leitoras a futuros movimentos de formação e articulação profissional.

Esperamos que esta obra contribua para a disseminação do debate e do conhecimento produzido no seminário **Avaliação para o Investimento Social Privado: Metodologias**, fortalecendo ainda mais as iniciativas sociais brasileiras. Boa leitura!

Avaliação como aprendizagem

ANDRE DEGENSAJN

Diretor executivo do GIFE – Grupo de Institutos, Fundações e Empresas.

O lugar ocupado pela avaliação no investimento social privado brasileiro é um bom indicador do seu grau de maturidade, com as distintas perspectivas que caracterizam práticas das organizações do campo social, sejam financiadoras, executoras de projetos sejam combinações desses elementos. Compreender as origens, os movimentos e as tendências dessas perspectivas é fundamental para o árduo exercício de moldar a cultura de avaliação no Brasil.

Fundado em 1995, o GIFE reúne organizações de origem privada que investem seus recursos em ações de interesse público. Diferentemente de outros países, que adotaram o termo *filantropia*, o GIFE optou em se organizar em torno do conceito de investimento social privado. Essa decisão se deu em função de o conceito de filantropia, no Brasil, estar mais associado à caridade e ao assistencialismo do que à atuação de organizações que procediam, em sua maioria, do campo empresarial em busca de transformações positivas no quadro social do País.

Nesse deslocamento já se coloca uma questão importante para pensar avaliação. No campo social, um primeiro desafio a ser ultrapassado é o das boas intenções. Essa ideia está presente nas ações voluntárias tradicionais e, em alguma medida, também disseminada entre investidores sociais e organizações da sociedade civil. Ela reforça a ênfase na *ação* e não na *transformação* que é capaz de gerar, daí derivando um conjunto de visões e de práticas presentes nas avaliações. Se o recurso é privado e a intenção é

contribuir para o bem comum, por que se preocupar em avaliar? Não estamos todos remando na mesma direção? Ao final do dia, não estaremos todos contribuindo com causas sociais relevantes?

Para refletir sobre o lugar da avaliação no investimento social, é importante retomar o sentido da atuação de organizações privadas na direção do interesse público, que se justifica pela capacidade de gerar benefícios tangíveis para a sociedade. Nesse sentido é que se admite, por exemplo, que o Estado favoreça tal atuação por meio de um conjunto de incentivos fiscais, reconhecendo a legitimidade de organizações privadas em suas contribuições para o bem comum. No entanto, tal legitimidade precisa estar ancorada em evidências de que se está gerando, de fato, transformações relevantes para a sociedade. E, para tanto, não bastam boas intenções.

A construção do campo do investimentos social no Brasil deriva da confluência de uma tradição filantrópica, pouco orientada pela mensuração de impacto, com a prática empresarial, habituada a métricas de monitoramento e de avaliação e, principalmente, a gerar resultados de curto prazo. Aqui se situa um desafio bastante relevante: sabe-se que o tempo das transformações sociais não é o mesmo tempo dos negócios.

O campo social no Brasil é composto por organizações muito novas. Segundo dados da Fasfil 2010¹, que analisa as fundações privadas e associações sem fins lucrativos num universo mais amplo do que o setor de investimento social privado, a média de idade das organizações é 14 anos. Entre os associados do GIFE, 39% têm menos de 10 anos, e 62%, até 20 anos. Esse é um contexto que precisa ser considerado ao refletirmos sobre avaliação, que tende a trabalhar com ciclos longos quando pretende identificar os impactos alcançados.

É importante considerar que a avaliação não é um elemento isolado na gestão das organizações. Organizações com baixa institucionalização, planejamentos pouco consistentes e investimen-

1. **As Fundações Privadas e Associações sem Fins Lucrativos no Brasil – 2010.** IBGE, 2012. Disponível em: <<http://www.ibge.gov.br/home/estatistica/economia/fasfil/2010/>> (último acesso em setembro de 2013).

tos descontínuos são pouco propensas a desenvolver processos avaliativos efetivos. Ou seja, a qualidade da avaliação é causa e efeito do fortalecimento e maturidade institucionais. Visão estratégica de longo prazo e estruturas de governança que deem consistência e continuidade ao investimento são fundamentais para a construção de avaliações melhores.

Nos últimos anos, no entanto, houve avanços expressivos na compreensão sobre avaliação, que se deslocou para um espaço mais central na atuação de investidores sociais. É indicativo desse movimento que 96% dos associados do GIFE declarem realizar avaliação de resultados². Pode-se dizer que uma cultura de avaliação está instaurada e que há um amplo reconhecimento de sua importância na qualificação dos investimentos. Apesar de essa constatação ser um passo fundamental, não se pode concluir que seja uma indicação direta da qualidade e da consistência das avaliações.

Como esse campo é marcado mais por diferenças do que semelhanças, convivem ainda visões contraditórias. Persiste a visão de que avaliar é perda de tempo ou de recursos, ou ainda que avaliação é apenas um instrumento de prestação de contas³. O estereótipo das boas intenções retorna: *Sabemos que estamos fazendo um bom trabalho; a avaliação é apenas funcional, para provar isso aos outros.*

Essas percepções sobre avaliação não são exclusividade de organizações no Brasil. Apesar da tentação de acreditar que “lá fora” as coisas são bem diferentes, um estudo conduzido pelo European Foundation Center com 56 fundações europeias, indica desafios muito semelhantes aos nossos⁴. Metade das organizações entrevistadas não acredita que seu trabalho, orientado a provocar transformações sociais, pode ser avaliado de forma consistente. Persiste a visão de que a avaliação é uma prática burocrática que impede

2. **Censo GIFE 2011-2012.** GIFE, 2012. Disponível em: <http://www.gife.org.br/publicacao-censo-gife-20112012-d23543a904f46c81.asp> <último acesso em setembro de 2013>.

3. Cf. pesquisa **Abordagens e práticas da avaliação de iniciativas sociais no Brasil.** Instituto Fonte, Fundação Itaú Social, Instituto Paulo Montenegro, Ibope Inteligência, 2012.

4. Cf. **Ceci... n'est pas un evaluation guide!** European Foundation Center, 2013. Disponível em: <http://www.efc.be/programmes_services/resources/Documents/Learning_Lab5.pdf> <último acesso em setembro de 2013>.

a inovação e a tomada de risco. Muitos acreditam que podem avaliar instintivamente o impacto alcançado. Em outro sentido, visões simplificadoras da avaliação convivem com esse ceticismo, reforçando a crença em modelos *one size fits all*.

Mas, como seria de esperar, há visões que apontam para a avaliação como um processo constante de aprendizado. Ou, mais do que isso, que exigem a construção de uma *cultura de aprendizagem* nas organizações. É nesse campo que se insere um dos principais desafios para a qualificação da avaliação no Brasil: como transformar a avaliação em um instrumento estratégico de aperfeiçoamento, de aprendizado, de redefinição e de reflexão sobre a atuação de investidores sociais?

Sem a pretensão de responder de maneira exaustiva a essa pergunta, um dos desafios que se coloca é o que poderíamos chamar de *cultura do êxito*. É sintomático que falemos tanto em *desafios*, mas raramente em *problemas* ou *dificuldades*. A apropriação da avaliação para a prestação de contas ou, ainda pior, para o *marketing social*, dificulta o reconhecimento daquilo que não funcionou e limita o espaço para o aperfeiçoamento e compartilhamento de experiências, não apenas as exitosas, mas também as que não foram capazes de gerar os resultados almejados.

A cultura do êxito favorece a visão de julgamento sobre as ações, estabelecendo dicotomias de sucesso e fracasso. Avanços efetivos no campo social se dão pela sedimentação de experiências, mais ou menos bem-sucedidas, capazes de incorporar erros e acertos em seu processo de construção. Ao contrário de inibir, esse reconhecimento tende a favorecer a inovação e a capacidade de correr riscos. E é fundamental recordar que essas são potencialidades essenciais do investimento social privado.

Ao observar dados do Censo GIFE sobre avaliação, é possível fazer uma curiosa inferência: quanto “piores” os dados sobre avaliação, melhor! Em 2010, 80% dos respondentes declaravam realizar avaliação de impacto em um ou mais projetos. Em 2012, esse número caiu para 54%, redução não compreendida como retrocesso nas práticas, mas como ampliação do senso crítico sobre os significados da avaliação de impacto. Assim como não se pode defender que atualmente mais da metade dos associados do GIFE

realiza, de fato, avaliações de impacto, seria também insensato defender que em 2010 a situação fosse melhor.

Há um reconhecimento de que houve avanços nessa área, mas ainda há uma baixa compreensão sobre as distinções entre as estratégias avaliativas disponíveis e o baixo aproveitamento das avaliações. Esse contexto abre espaço para a construção de uma visão setorial sobre a avaliação e de referenciais comuns que auxiliem no desenvolvimento de uma cultura de aprendizagem mais profunda no setor.

A ampliação da compreensão sobre as abordagens metodológicas e sobre a efetiva incorporação da avaliação nas estratégias de investimento tende a favorecer o reconhecimento do seu valor em diferentes dimensões: em termos de retorno econômico do investimento social; da justa e efetiva aplicação de recursos que se beneficiaram de algum incentivo fiscal; e da percepção pública do valor e da contribuição dos investidores sociais ao desenvolvimento do País.

A escolha de métodos para fazer avaliação

ROGÉRIO SILVA E DANIEL BRANDÃO

Sócios-diretores da Move Avaliação e Estratégia em Desenvolvimento Social.

1. INTRODUÇÃO

À medida que as organizações da sociedade civil brasileiras incrementam o uso estratégico de processos avaliativos, certas inquietações têm crescido entre gestores e avaliadores, num evidente sinal do amadurecimento das práticas e do campo. Sendo parte desse movimento, é provável que o leitor familiar ao *mundo das avaliações*, seja contratando-as, conduzindo-as ou delas tomando parte, já tenha se debruçado sobre perguntas tais como: “Qual método é mais adequado para avaliar este ou aquele programa?”; “Qual tipo de abordagem é mais precisa?”; “Que tipo de informação é mais relevante e convincente?”; “Que abordagem pode ser mais profunda e sensível à realidade social?”; “Que estratégia metodológica é mais abrangente?”.

Este ensaio tem o propósito de introduzir elementos capazes de auxiliar na resposta a perguntas como essas. Lançando mão de reflexões teóricas articuladas ao pragmatismo de casos em que atuamos nos últimos anos, projetamos a escolha dos métodos como função de seis variáveis-chave às quais daremos atenção a seguir. Longe de esgotá-las, papel realizado por outros artigos deste livro e por outras publicações, propomos abordar as escolhas metodológicas conforme expresso na figura a seguir, reconhecendo as características dinâmica e multidimensional da escolha de métodos. >

O ATO MULTIDIMENSIONAL E DINÂMICO DA ESCOLHA METODOLÓGICA



2. CONSIDERAÇÕES SOBRE OS CAMINHOS METODOLÓGICOS

O debate sobre métodos ocupa lugar privilegiado no campo da avaliação, a ponto de alguns autores tomarem a maneira de investigar a realidade como único critério de julgamento válido para definir a qualidade de uma avaliação. Não é à toa que inúmeras publicações sejam inteiramente dedicadas às origens e sentidos das raízes metodológicas da avaliação contemporânea. Em casos extremos, vê-se o debate metodológico reduzido à escolha estritamente científica de técnicas e instrumentos de medida, pondo à margem quaisquer outras conexões entre a avaliação e as realidades socioinstitucionais em que ocorrem e, em particular, a relação umbilical entre avaliação e julgamento.

Marvin Alkin (2004), em seu clássico *Evaluation roots*, foi um dos que primeiro reconheceu que o debate sobre métodos esta-

va pressionado por aqueles que achavam que apenas os estudos experimentais combinando grupos de controle e de intervenção, os famosos Randomized Control Trials (RCT), seriam capazes de produzir avaliações de verdade, com incomparável capacidade de atribuir causalidade. Nesse caso, o autor chamava atenção para o fenômeno de captura do julgamento por uma onda de rigor científico. Curiosamente, esse debate atravessou décadas, estando atualmente presente também entre gestores e avaliadores brasileiros.

Em paralelo, muitos autores têm advogado por um encontro mais dialógico entre rigor e uso, fazendo do debate metodológico um campo em que posições ideológicas e técnicas, em que sujeitos e dinâmicas organizacionais são tomados como reguladores dos métodos, assim submetidos ao jogo democrático. Entre os autores mais proeminentes dessa linha, Guba e Lincoln (2000) irão filiar as escolhas metodológicas ao posicionamento dos sujeitos frente aos paradigmas científicos, localizando os métodos como realizações práticas de posições científicas e políticas bem mais profundas. Para eles, não haveria escolha metodológica fora dos quatro paradigmas científicos hegemônicos – positivismo, pós-positivismo, teorias críticas e construtivismo –, sendo necessário reconhecer onde se está, no que se acredita e o que se deseja, para fazer escolhas coerentes.

Em leituras mais pragmáticas, Ernest House (1983) reconhecia, já no início dos anos 1980, diferentes abordagens práticas que caracterizavam as avaliações. A classificação proposta pelo autor, retomada por vários outros, ofereceu a possibilidade de eleger modelos com base na capacidade dos estudos em: (1) medir o alcance dos objetivos planejados; (2) responder às perguntas dos principais gestores da iniciativa; (3) auxiliar usuários e consumidores a elegerem serviços e produtos; (4) lançar mão de critérios técnico-científicos para julgar a realidade; (5) favorecer a comparação entre visões antagônicas de uma mesma iniciativa; e, finalmente, (6) abrigar as perspectivas de diferentes interessados no objeto avaliado.

Quaisquer que sejam as portas de entrada para a escolha de métodos, o que os anos de prática no campo social nos mostraram é a inadequação das posições que insistem em reivindicar para si o “cetro real do método sagrado”, um tipo de solicitação que mais re-

memora o narcisismo perdido dos primeiros anos de vida do que evoca o estatuto fundamental da ciência contemporânea: compreender, enxergar e praticar ciência como debate, como enuncia Pedro Demo (DEMO, 2011).

Em boa medida, escolher métodos demanda compreendê-los como produtos de um itinerário essencialmente ético e político. Ético, se reconhecermos a importância de submeter os métodos à ordem da realidade, para colocá-los a serviço de um problema experimentado por sujeitos concretos, para construí-los em resposta a uma questão de investigação, modulando-os em função do contexto sócio-histórico no qual estão inseridos. Político, se apossarmos que tais escolhas se dão em ambientes essencialmente coletivos e plurais onde sempre se está intervindo, produzindo e transformando, nos lembrando de que toda avaliação está inserida e é função de um ecossistema particular.

O pragmatismo que cultivamos como *practitioners*, no melhor sentido da palavra, nos faz afirmar que métodos não são avaliações, ainda que sejam determinantes na modelagem de suas características e utilidade. Transformar um método na principal preocupação de um gestor amplia o risco de esvaziar o potencial estratégico da avaliação, reduz a dinâmica de aprendizagem e compromete a relação entre avaliação e desenvolvimento organizacional. Informações rigorosas não são suficientes nas trincheiras do mundo real. Opor técnica e plástica, ciência e política, saber e estética são armadilhas a serem constantemente evitadas.

Vejamos as variáveis-chave que compõem a espinha dorsal deste ensaio: o ato multidimensional e dinâmico das escolhas metodológicas. Multidimensional porque diferentes variáveis compõem o cenário. Dinâmico porque as relações entre elas, bem como sua participação nas escolhas, estão em constante movimento.

3. VARIÁVEIS PARA A ESCOLHA DOS MÉTODOS

3.1. NATUREZA DO OBJETO DE ESTUDO

Deve-se reconhecer o objeto de estudo para compreender as características que serão fundamentais na eleição do método. Se o

objeto é do campo A ou B, como a Educação Básica, o empreendedorismo ou a agricultura familiar, será preciso acessar as bases teóricas, políticas e normativas que o modulam e que conferem seu *status* na contemporaneidade das ciências e das articulações sociais que o fundamentam. Por exemplo, como estudar um programa que visa qualificar práticas de Educação Infantil sem levar em conta as diretrizes curriculares nacionais para esta área?

Nessa aproximação, será fundamental compreender também as expectativas que cercam um determinado projeto ou programa, as relações lógicas que estabelecem entre insumos, processos e resultados ou sua teoria de mudança. Por exemplo, como avaliar um projeto que visa renovar o estoque de fauna de um determinado conjunto de lagoas, sem levar em conta que centros de pesquisa, associações de pescadores, órgãos ambientais e um comitê de bacia possuem papéis complementares entre si, num processo técnico e político que só pode esperar impactos num intervalo de dez anos?

Dessa forma, o *status* teórico, político e normativo de um objeto, somado aos atores – o que compreende e envolve as relações, os recursos que utiliza, os resultados que espera e a sua lógica de intervenção –, são elementos que não podem ser esquecidos. Fechar os olhos para eles supõe que um método pode ser escolhido à revelia do objeto, o que tende a levar a leituras superficiais e a julgamentos impróprios.

3.2. POSIÇÕES TEÓRICAS DOS INTERESSADOS

À natureza do objeto articulam-se as posições e preferências teóricas dos principais interessados em uma avaliação. Partindo do princípio de que o conjunto de interessados deve legitimar a avaliação para utilizá-la, é fundamental que qualquer estudo seja capaz de produzir saberes inteligíveis e confiáveis a quem ele se destina. Se um programa voltado a incrementar a renda de um conjunto de mulheres tem como seus principais interessados um grupo de executivos do mercado financeiro, cuja familiaridade com teorias econômicas e com a linguagem matemática é elevada, haveria motivo para não dar ênfase a uma abordagem econométrica?

Por outro lado, se um programa com a mesma finalidade tem como principais interessados um conselho composto por especialistas na multidimensionalidade da pobreza, por que não dar ênfase a um método que privilegie o estudo de casos a partir de entrevistas em profundidade com sujeitos tipo ideal, culminando em análises de conteúdo críticas em sua capacidade de explicitar resultados e de marcar contradições e limites da intervenção?

Na tentativa de evitar rótulos, mas também de tornar plano um terreno de posições teóricas que tende a ser bastante irregular, é fundamental que as escolhas metodológicas acessem as posições e preferências teóricas dos interessados, seja para que sua oferta de resultados responda a elas em harmonia, seja para que se escolha um caminho consciente para contrapor tais posicionamentos e provocar os interessados a expandirem suas compreensões.

3.3. CULTURA ORGANIZACIONAL

Se a natureza do objeto é capaz de refutar certos caminhos e apontar outros, e as posições teóricas dos interessados ajudam a delinear caminhos analíticos e, em certa medida, a linguagem da avaliação, a cultura organizacional em que um estudo se dá é um componente determinante do sentido e do potencial de uma avaliação.

Em texto anterior, apontamos quão delicadas são as relações entre avaliação e cultura organizacional, destacando que quanto menos os estudos são tratados como próteses e mais como órgãos, mais ampliam sua potência transformadora nas organizações (SILVA, 2012). Nesse sentido, a eleição de métodos não pode se dar sem que certa leitura da organização seja realizada, e que essa leitura contribua em sua modulação.

De que maneira a organização faz e usa planejamento? Que processos de aprendizagem se dão ali? De que maneira a organização lida com erros e acertos, fracassos e sucessos? Há conflitos em torno do projeto ou programa que será avaliado? De que maneira a organização costuma conversar sobre seus projetos, programas e estratégias? Outras avaliações ali realizadas produziram que tipos de fenômenos? Leituras advindas de perguntas como essas podem ser de grande ajuda na eleição de métodos e até mes-

mo para que se possa reconhecer um momento inoportuno de se realizar uma avaliação.

Se os processos avaliativos reúnem a possibilidade de explicitar e revisar premissas, eventualmente expondo suas fragilidades; se eles são potencialmente estimuladores da reflexão e do diálogo investigativo, já que partem da irrefutável vontade de saber, acessar e manejar a permeabilidade da cultura organizacional aos processos avaliativos, é crucial eleger métodos dosando cuidado, respeito e prudência. Uma vez mais, se o caminho é o da contestação ou do uso da avaliação nas vezes de um “cavalo de Troia”, recomenda-se reflexão tática, para evitar heroísmos danosos.

3.4. VIABILIDADE ECONÔMICA

O volume de recursos financeiros, temporais e estruturais alocados em uma avaliação guarda proporção direta com o desenho metodológico utilizado. O mais complexo, volumoso e rigoroso exige também um investimento maior. Contudo, ao que se deve ficar muito atento, é que não se pode dizer o mesmo da relação entre investimentos e utilidade de uma avaliação. Há casos de intensa eficiência, bem como os de vergonhoso desperdício.

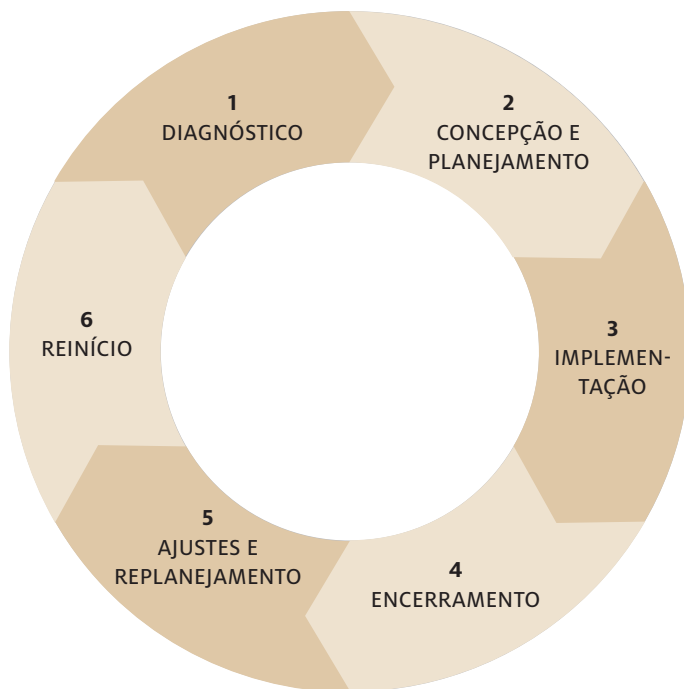
O equilíbrio entre o *preciso* e o *viável* deve orientar gestores e avaliadores na elaboração dos desenhos que serão adotados. Tendo em vista as dimensões, já abordadas neste artigo, que irão modular as escolhas, o potencial de investimento em uma avaliação deve ser considerado de maneira cuidadosa em sua capacidade de regular a precisão necessária, o tempo de execução, as demandas dos interessados e, por fim, a capacidade de permitir que se construam respostas para as perguntas em questão. Como a tarefa não é simples, elegemos alguns tópicos para aprofundar a discussão, no quadro a seguir. >

MITOS E REALIDADES DOS INVESTIMENTOS EM AVALIAÇÃO	
Mitos	Realidade apreendida no longo tempo de trabalho
Boas avaliações consomem de 5% a 10% do orçamento de um programa.	Avaliações brilhantes são realizadas com investimentos muito variáveis, sendo inadequado, senão impossível, propor uma associação entre investimento e qualidade. Quanto maior o volume de orçamento de um programa, menor será, percentualmente, o investimento em avaliação. A sugestão é que, ao se planejar um programa, se planeje também, ainda que em linhas gerais, seus processos de avaliação.
Estudos qualitativos são mais onerosos do que estudos quantitativos.	Três componentes são centrais nos orçamentos de uma avaliação: (a) o volume de tempo de trabalho de avaliadores seniores e/ou especialistas, geralmente destinados ao desenho e análises; (b) o volume de trabalho de pesquisadores e/ou avaliadores assistentes, geralmente destinados a coleta de dados e pré-análises; (c) os custos de logística, tais como passagens, hospedagem etc. Há muitos casos, sobretudo os estudos multicêntricos, em que estudos quantitativos consomem mais recursos do que os qualitativos.
Avaliações internas são menos onerosas do que avaliações externas.	Quando se leva em conta o desembolso direto, o mito se confirma. Por outro lado, se os custos de oportunidade forem olhados com cuidado, e somados à efetividade dos processos, muitas vezes avaliações externas custam menos.
Avaliações neutras só podem ser externas e desse investimento não é possível escapar.	As equipes internas têm seus interesses e limites, e podem também ser incentivadas a desenvolver competências avaliativas que confirmam precisão e equilíbrio aos estudos. Os avaliadores externos, se por um lado costumam gozar de maior isenção da dinâmica organizacional, também têm suas preferências, suas tendências e seus pontos cegos. A busca de um processo equilibrado em sua capacidade de enxergar luzes e sombras não pode terminar refém de avaliadores externos.
A oferta de serviços de avaliação é pequena no Brasil, tornando-os onerosos.	O cenário brasileiro oferece centenas de profissionais autônomos, universidades, organizações não governamentais prestadoras de serviços e empresas especializadas em diferentes tipos de avaliação. Em função de seus diferentes arranjos institucionais e sua distribuição em diferentes regiões do Brasil, seus orçamentos são extremamente variantes e viáveis para organizações de diferentes portes.

3.5. ENQUADRE GERENCIAL

Ao lado das dimensões até aqui explicitadas, cabe adicionar a perspectiva do ciclo de gestão de um projeto ou programa, também essencial à eleição de métodos. Em ciclos de gestão tradicionais, como o retratado na figura a seguir, cada um dos momentos tende a produzir perguntas com ênfases distintas, a serem exploradas também de maneiras distintas. >

CICLO DE GESTÃO DE UM PROJETO OU PROGRAMA SOCIAL



Nos momentos que antecedem a implementação, é de se supor que perguntas de duas naturezas tendam a emergir. A primeira delas diz respeito aos resultados que poderão ser alcançados pela iniciativa, enquanto a segunda, ao desejo de comparar os resultados futuros com a situação presente. Essa dinâmica tende a produzir o que se costuma chamar de estudos de linha de base, que irão eleger e medir indicadores a serem tomados como base para futuras comparações.

Ao longo dos processos de implementação, outros três tipos de pergunta podem ganhar força e ajudar a moldar as escolhas metodológicas – indagações relacionadas à dinâmica da equipe de trabalho, às relações entre equipe e beneficiários, à capacidade de aprender e se ajustar à realidade, entre outras. Ao mesmo tempo, podem haver questões concentradas em compreender em que medida as metas planejadas (ou resultados parciais) estão sendo alcançadas, remetendo o processo ao *status* de um monitoramento. Por fim, pode haver momentos em que avaliações de meio-termo

sejam necessárias, ou seja, aquelas que procuram articular os resultados parcialmente alcançados à qualidade dos processos de trabalho, propondo revisões no planejamento inicial e repactuação dos resultados esperados.

Nos períodos de encerramento, são as avaliações de resultados as que ganharão força, sendo essencialmente traduzidas pelo julgamento da relevância e do mérito com os quais a iniciativa alcançou o final de seu ciclo. Nesse momento, a existência de registros de linha de base pode ser um elemento importante para definir o desenho do estudo. Contudo, a ausência de um levantamento dessa natureza não inviabiliza a comparação, mas demandará técnicas retrospectivas para comparar presente e passado.

É também a partir da finalização de um ciclo de projeto que se pode pensar em perguntas de impacto, como aquelas que tentarão identificar efeitos e consequências socioambientais, econômicos, culturais e políticos de longo prazo, previstos e imprevistos, negativos ou positivos, alcançados a partir de uma intervenção em um determinado grupo populacional. Sustentabilidade, profundidade, influência, escala e abrangência são significantes que tendem a emergir nesse momento, num desejo de atribuição de causalidade entre iniciativa e transformação social.

3.6. POSIÇÕES ÉTICO-POLÍTICAS

O estudo de situações sociais implica, inexoravelmente, a observação da relação de homens e mulheres entre si e com o mundo ou o território que habitam. A dinâmica estabelecida entre os diversos sujeitos participantes de uma realidade social é atravessada por relações de interesse e poder que podem, ou não, se expressar por meio das opções metodológicas presentes em uma avaliação.

O método cria configurações que podem fortalecer determinados indivíduos e enfraquecer outros, seja pela sua possibilidade de ser compreendido por diferentes interlocutores seja por permitir envolvimento e participação em decisões relevantes. Da mesma forma, o método pode privilegiar o lugar social de um certo gênero, de uma classe social, de um *status* político, e assim por diante, amplificando uns e calando outros, ainda que o faça sem consciência.

Ao lado das outras cinco dimensões apresentadas, as posições ético-políticas de gestores e avaliadores terão grande peso na determinação dos métodos avaliativos. A quem pertence um processo de avaliação? Quem deve e não deve falar ao longo do processo? Que perspectivas devem ser priorizadas? Quem toma parte e de que maneira? O encontro com tais perguntas é um componente do qual não se deveria fugir.

4. CONSIDERAÇÕES FINAIS

Método é palavra a ser pensada no plural: *métodos*. Não existe um método, tampouco “o método”, mas inúmeros arranjos essencialmente caracterizados pelas correntes epistemológicas que os fundamentam, que abrigam não apenas um caminho técnico de acesso à realidade, mas um caminho mediado por sujeitos que pensam e aspiram, num tabuleiro com tantos outros desejos e ideologias.

Distante do pragmatismo necessário aos processos de avaliação, um método não é nada. Ele assume seu lugar à medida que se propõe a responder perguntas cruciais para uma iniciativa, num esforço legítimo para compreender a realidade e, dessa maneira, alterá-la. Quando articulado no campo avaliativo, um método é um dispositivo a serviço de investigar, compreender e ajudar a mudar. Como versa a origem grega da palavra, com *meta* (relacionado a resultado) e *hódos* (caminho), os métodos são caminhos, estradas para produzir resultados.

Nesse sentido, é fundamental que as escolhas metodológicas não sejam tomadas por gestores e avaliadores como a preocupação central do processo avaliativo. Fetichizado, carregado de atributos e expectativas que não correspondem ao que ele é ou ao papel que deve cumprir, abre-se uma forte via para a frustração e o engodo.

Mais ainda, é essencial que não se construa nem se alimente fantasias de que os métodos estão encastelados em torres acessíveis apenas a especialistas, como se fossem cavaleiros míticos dotados do poder de vencer o dragão. O debate pode e deve ser democrático, acessível, disponível, num esforço de elevar a capacidade dos sujeitos para melhor compreenderem o seu entorno, numa cartografia que inclua as pedras e os espectros.

BIBLIOGRAFIA

- ALKIN, M. C. (Ed.). **Evaluation roots**: tracing theorists' views and influences. Thousand Oaks, CA: Sage, 2004.
- DEMO, P. **Praticar ciência**. Metodologias do conhecimento científico. São Paulo: Saraiva, 2011.
- GUBA, E. G.; LINCOLN, Y. S. Paradigmatic controversies, contradictions, and emerging confluences. In: DENSZIN, N.; LINCOLN, Y. S. (Eds.). **Handbook of qualitative research**. 2. ed., Thousand Oaks: Sage Publication, 2000, pp. 163-88.
- HOUSE, E. R. Assumptions underlying evaluation models. In: MADDAUS, G. F.; SCRIVEN, M.; STUFFLEBEAM, D. L. (Orgs.). **Evaluation models**: viewpoints of educational and human services evaluation. Boston: Kluwer-Nijhoff, 1983.
- SILVA, R. R. Avaliação e desenvolvimento organizacional. In: OTERO, M. R. (Org.). **Contexto e prática de avaliação de iniciativas sociais no Brasil**: temas atuais. São Paulo: Peirópolis/Instituto Fonte, 2012.
- SILVA, R. R. Avaliação de programas e projetos sociais no Brasil: história, conquistas e desafios. In: **Avaliação de programas e projetos sociais no Brasil**. A experiência da Fundação Banco do Brasil. Brasília: FBB, 2013

Métodos mistos de avaliação: aumentando a efetividade do investimento social privado

DONNA M. MERTENS

PhD, Universidade de Gallaudet, Washington, DC, Estados Unidos.

1. INTRODUÇÃO

O investimento social privado tem grande potencial para melhorar as condições de vida dos povos mais pobres e destituídos do mundo. No Brasil é evidente que a Fundação Itaú Social, Fundação Roberto Marinho e Fundação Maria Cecília Souto Vidigal, entre outras, oferecem um volumoso e importante suporte. Mas a consciência sobre a responsabilidade social inclui não apenas a oferta de fundos e programas, mas também o trabalho de avaliar rigorosamente as iniciativas, a fim de responder às demandas das diferentes comunidades alcançadas pelo investimento.

Em diferentes setores, o uso da avaliação aumentou significativamente nas duas últimas décadas. Na arena política, os atores responsáveis pela formulação e implementação de programas e serviços – especialmente em uma época de turbulência econômica e austeridade fiscal – são constantemente pressionados por seu eleitorado acerca dos resultados e do retorno econômico do investimento público. Também no contexto dos programas sociais, as

organizações não governamentais avaliam frequentemente a eficácia de suas estratégias de implementação, para aprender como melhor servir a seus beneficiários e assegurar seu financiamento. Além disso, gestores de iniciativas do setor privado usam avaliação para entender quais são as correções necessárias para melhorar suas estratégias de marketing e vendas e aumentar a satisfação de seus clientes, além de aumentar lucros.

Abordagens de avaliação que usam métodos mistos fortalecem a habilidade da comunidade avaliativa em responder a essas demandas com maior rigor, responsabilidade e eficiência em grupos culturalmente diversos, tais como os encontrados no Brasil. Ao mesmo tempo em que essa comunidade endossa o uso de métodos mistos com esse fim, não há um consenso sobre como integrá-los em seu trabalho. Entre os avaliadores as divergências que justificam o uso das combinações de métodos têm base nas premissas que guiam a conduta de cada profissional.

Há quatro abordagens principais para o uso de métodos mistos, que enfatizam diferentemente o impacto, o processo, a relevância cultural e o uso das avaliações (MERTENS & WILSON, 2012; MERTENS & HESSE-BIBER, 2013). Para aqueles que acreditam que há uma realidade a ser medida, não importando o quão imperfeita, o enfoque está na avaliação do impacto e no uso de métodos quantitativos, com um componente qualitativo. Outros avaliadores, construtivistas, acreditam que há várias realidades tecidas socialmente e que os profissionais precisam interagir com as comunidades para poder relatar a natureza da situação vivida por seus integrantes. Assim sendo, priorizam o uso de métodos qualitativos, com a inclusão de dados quantitativos como parte das evidências.

Um terceiro grupo, formado pelos pragmáticos, afirma que as perguntas avaliativas vão determinar a escolha dos métodos e que, na maioria dos casos, a compreensão será melhor lançando mão de métodos qualitativos e quantitativos. O quarto grupo, de postura mais transformadora, enxerga a avaliação como uma maneira de contribuir para mudanças sociais positivas. Dito isso, o uso de métodos mistos será necessário para entender a complexidade de comunidades culturalmente diversas e garantir que os direitos humanos e a justiça social sejam priorizados. Avaliadores que

endossam o pluralismo dialético (JOHNSON & STEFURAK, 2013) recomendam que se utilizem as premissas filosóficas do paradigma construtivista em seu trabalho qualitativo e as do paradigma pós-positivista em seu trabalho quantitativo. A mistura começa na hora de relacionar e cruzar os dados coletados segundo essas diferentes posições paradigmáticas.

Essas posturas implicam na forma como tais métodos mistos são usados em avaliação. Os profissionais que se alinham ao paradigma pós-positivista enfatizam o uso de estudos randomizados controlados (RCT, na sigla em inglês) como a melhor maneira de responder às perguntas voltadas a estudar o impacto e a eficácia de um programa (WHITE, 2013). Eles reconhecem o valor de se acrescentar um componente qualitativo na etapa de coleta de dados sobre as “cadeias causais, a escolha de público-alvo, as barreiras para a participação e a adesão dos beneficiários” (p. 62).

Já os construtivistas defendem a inclusão de “perspectivas metodológicas subjetivas ligadas a uma série de métodos cuja ontologia favorece uma visão da realidade como múltipla, tendo o propósito de entender múltiplas subjetividades” (HESSE-BIBER, 2013, p. 51). Isso permite a coleta e o uso de dados qualitativos antes, durante e depois que uma intervenção é implementada, produzindo respostas para perguntas como: “O quanto essa intervenção responde à cultura e ao contexto da população-alvo?”; “Até que ponto a população-alvo dialoga com a diversidade de objetivos do projeto?”; “Quem ficou de fora?”; “Por quê?” (p. 52). A postura do pluralismo dialético defende o uso de avaliadores que partem de premissas filosóficas divergentes para proteger a integridade das decisões metodológicas e mantê-las de acordo com as crenças de cada paradigma, ao mesmo tempo em que permite a conversa frutífera que surge quando diferentes perspectivas são respeitadas.

Nem o paradigma pragmático nem o transformador preconizam o uso de métodos mistos. Mas o pragmatismo tem sido usado como base para defender a utilização de métodos mistos quando as perguntas avaliativas podem ser mais bem respondidas com essa abordagem. Em contraste, o grupo de avaliadores com postura transformadora questiona a premissa de que as perguntas de uma avaliação são predeterminadas e apoiam o envolvimen-

to com todos os interessados de modo culturalmente respeitoso (MERTENS, 2009; MERTENS & WILSON, 2012). Isso significa que os avaliadores têm de fazer algum trabalho qualitativo preliminar para entender a comunidade antes de desenvolver as questões da avaliação, de modo a construir uma relação com ela e garantir que os grupos interessados sejam incluídos e apoiados.

É necessário identificar a natureza das relações de poder e suas implicações sobre as vozes que deverão ser privilegiadas no desenho da avaliação e do programa. Os métodos devem ser capazes de identificar as versões da realidade sustentadas por cada grupo e compreender as implicações dessas versões na busca por justiça social e, em última instância, por direitos humanos. Por exemplo: se um grupo acredita que a razão para o aumento da Aids/HIV no Hemisfério Sul deve-se à ignorância da população sobre como evitar o contágio, pode propor uma intervenção ensinando as pessoas a “simplesmente dizer não” ou usar preservativos. Porém, essa intervenção não parece baseada na compreensão e/ou experiências com Aids/HIV entre as pessoas. Uma avaliação transformadora usaria dados quantitativos e qualitativos coletados em entrevistas, observações e convívio com o público-alvo, além de dados epidemiológicos desagregados em dimensões relevantes para compreender o sentido da questão entre homens e mulheres (diversidade), o efeito da pobreza na disposição em praticar sexo sem proteção, a sensação de futilidade sobre o futuro quando tantas pessoas à volta estão morrendo, e outros fatores que precisam ser incorporados aos planos para se chegar a uma intervenção capaz de provocar mudanças (CHILISA, 2012).

Assim sendo, os métodos mistos têm grande potencial para aumentar a credibilidade das evidências que os avaliadores podem oferecer às comunidades para apoiar a efetividade, propriedade e sustentabilidade dos investimentos sociais privados no Brasil. Seja trabalhando com educação, saúde, desenvolvimento social e econômico, melhorias de infraestrutura, setores ambientais ou artes e cultura, basear o trabalho no claro entendimento das experiências das pessoas e alinhar a avaliação com as premissas filosóficas apropriadas aumentam a possibilidade de contribuir para melhorar a justiça social e garantir os direitos humanos.

O restante deste artigo traz uma análise detalhada dos diferentes paradigmas e tipos de avaliação, com exemplos que ilustram como os estudos com métodos mistos são conduzidos a partir dessas diferentes perspectivas.

2. MÉTODOS MISTOS E MÚLTIPLOS MÉTODOS EM AVALIAÇÃO: UMA BREVE HISTÓRIA

Após a publicação do primeiro enquadramento conceitual sobre avaliação com métodos mistos no fim dos anos 1980 (GREENE, CARACELLI & GRAHAM, 1989) e com base no aumento da consciência dos profissionais sobre a relevância de seu uso, ocorreu uma série de importantes avanços no mundo avaliativo, como se descreve a seguir.

- Uma edição inovadora do *New Directions in Evaluation* (Novas Diretrizes na Avaliação, em tradução livre) é dedicada aos métodos mistos (GREENE & CARACELLI, 1997), com o objetivo de listar os avanços alcançados com seu uso na avaliação no fim dos anos 1990.
- O *User-friendly handbook for mixed method evaluations* (Guia para o uso de métodos mistos em avaliação, em tradução livre) é publicado pela Fundação Nacional de Ciência (NSF na sigla em inglês, em 1997).
- São publicadas a primeira e a segunda edições do *Handbook of mixed methods in social and behavioral research* (Manual dos métodos mistos em pesquisa social e comportamental, em tradução livre) (TASHAKKORI & TEDDLIE (2003; 2012)).
- É lançada uma publicação revisada por avaliadores, especificamente dedicada ao uso de métodos mistos: o *Journal of Mixed Methods Research* (2007).
- O segundo volume do *New Directions in Evaluation* é dedicado ao uso de métodos mistos em 2013, focando em sua contribuição para a credibilidade das evidências (MERTENS & HESSE-BIBER, 2013).

Concomitante com a disseminação desses avanços, estabeleceram-se grupos profissionais e comunidades de prática com enfoque específico nos métodos mistos, incluindo os listados a seguir.

- O Special Interest Group (SIG na sigla em inglês, Grupo de Interesse Especial em métodos mistos) criado na American Educational Research Association (Associação Americana de Pesquisa Educacional), em 2009.
- O Topical Interest Group (TIG, Grupo de Interesse Temático) nos métodos mistos estabelecido na American Evaluation Association (AEA, Associação Americana de Avaliação), em 2010¹.
- O lançamento da Mixed Methods International Association (MMIRA, Associação Internacional de Métodos Mistos), em 2013.

E diante da crescente popularidade do tema, tanto dentro como fora dos meios acadêmicos, uma série de conferências sobre o uso de métodos mistos foi realizada no mundo, entre elas:

- Conferência Internacional de Métodos Mistos em Cambridge e Leeds (Reino Unido) e Baltimore, Maryland (Estados Unidos) em 2005-2012.
- O “Dia dos Métodos Mistos”, evento de um dia de aprendizado organizado antes da Conferência Internacional de Pesquisa Qualitativa na Universidade de Illinois, Urbana/Champaign (EUA), em 2012 e 2013.
- A conferência anual da MMIRA em Boston (EUA), em 2014.

Mais recentemente, as avaliações por métodos mistos receberam atenção especial em duas comunidades. Na área das ciências da saúde, o National Institutes of Health (NIH), agência do Departamento de Saúde e Serviços Humanos dos EUA, publicou um manual com as melhores práticas para pesquisas com métodos mistos (NIH, 2011). Na comunidade internacional, o Departamento para o Desenvolvimento Internacional britânico (DFID, 2012) publicou um artigo sobre como expandir a gama de formulações e métodos para ava-

1. Conforme a petição apresentada a favor da constituição da MM TIG dentro da AEA em 2010, a missão e o objetivo deste novo grupo são: (a) examinar o uso dos métodos mistos em avaliação por meio da análise reflexiva da filosofia, teoria e metodologia que está sendo desenvolvida no campo dos métodos mistos; (b) contribuir para a melhoria das práticas de avaliação, métodos e uso; e (c) focar nas contribuições que um melhor entendimento dos métodos mistos pode oferecer.

liações de impacto, e Bamberger (2012) publicou uma introdução ao uso de métodos mistos na avaliação de impacto. Mais precisamente, o interesse pela capacidade de os métodos mistos realizarem triangulação de dados e alcançarem melhor entendimento da implementação dos processos e causas dos fenômenos em estudo foi aguçado por dois fatores. Primeiro, pela identificação das limitações e desafios práticos associados aos RCTs, financiados por uma série de doadores internacionais; e segundo, a consciência da distinção entre qualitativo e quantitativo deixou de ser aparente devido à gama de ferramentas de coleta e análise de dados disponíveis para os avaliadores, como confirmado pelos autores da publicação da DFID:

A combinação de métodos também se tornou mais fácil porque a clara distinção entre métodos quantitativos (variáveis) e qualitativos (casos) se tornou mais diluída, com métodos quantitativos que não são estatísticos e novas formas de análises de casos, tornadas mais fáceis pelo uso de ferramentas eletrônicas (DFID, 2012, ii).

3. RELEVÂNCIA DA AVALIAÇÃO NA TEORIA E NA PRÁTICA

Como já foi mencionado anteriormente neste artigo, o eclético uso dos métodos mistos em avaliação está estreitamente ligado à variedade de propósitos aos quais se aplica a combinação de métodos quantitativos e qualitativos. Como resultado, os estudos baseados em métodos mistos confirmam a diversidade de estruturas filosóficas, lentes teóricas, escolhas metodológicas e práticas encontradas na comunidade de avaliadores. No campo dos métodos mistos, existem múltiplas perspectivas do que representa uma evidência ou do que dá vigor aos achados, estando estas também baseadas no nível de participação dos interessados, além de nas matrizes disciplinares dos avaliadores e nas escolhas de paradigma. Por esse motivo, os métodos mistos atuam em vários níveis.

Os diálogos sobre sua aplicação estão inerentemente ligados à credibilidade da evidência e ao critério usado para estabelecer tal credibilidade (MERTENS & HESSE-BIBER, 2013). Os avaliadores voltam-se para os métodos dos paradigmas pós-positivista, construtivista, pragmático e transformador para critérios associados

com credibilidade e rigor particulares a estes paradigmas (MERTENS & WILSON, 2012). Mas, longe de se preocupar somente com a solidez dos dados nos quais suas conclusões se baseiam, aqueles que usam os métodos mistos procuram representar da maneira mais fiel possível as vozes dos interessados menos favorecidos, de comunidades marginalizadas cujos pontos de vista vêm sendo historicamente ignorados ou representados de maneira pouco fiel nas avaliações. Em resposta à pergunta que eles normalmente se fazem, sobre como a inclusão dessas comunidades marginalizadas na avaliação aumenta a credibilidade das conclusões, defensores dos métodos mistos penam para justificar suas escolhas metodológicas, por causa da quantidade de opções. Tamanha diversidade e riqueza de metodologias apresenta alguns desafios, inclusive a necessidade de trabalhar dentro e entre dois pontos ou perspectivas de mundo que estão em tensão um com o outro.

4. O PARADOXO DOS MÉTODOS MISTOS

Como resultado do crescimento da literatura especializada na última década e da proliferação dos grupos profissionais citados anteriormente, o uso de métodos mistos passou a ser um aspecto distinto de muitas metodologias de avaliação (MERTENS & HESSE-BIBER, 2013). O esforço para combinar métodos quantitativos e qualitativos para aumentar a validade e credibilidade dos resultados representa um trabalho admirável. No entanto, seu uso não sistemático e não discutido também lança dúvidas sobre a eficácia de sua aplicação. Portanto, é necessária uma compreensão mais sólida e compartilhada dos métodos mistos. A seguir, apresentamos razões para seu uso².

2. Apesar do aumento, de modo geral, da aplicação de métodos mistos nas avaliações, eles ainda correspondem a uma pequena porcentagem do total de estudos conduzidos em uma série de campos. Um recente panorama de estudos publicados no *Journal of Organizational Behaviour* (2003-2008) citou a ocorrência em 17 estudos, de um total de 318. Uma revisão do *Journal of Addictive Behaviors* citou seis casos em 932 artigos, e um levantamento dos estudos sobre saúde na base de dados do *Computer Retrieval of Information on Scientific Projects* (CRISP, <<http://crisp.cit.nih.gov/>>, último acesso em setembro de 2013) cita apenas 272 entre 1997 e 2008. Ivankowa e Kawamura, 2010, observam que o número de estudos de métodos mistos cresceu de 10 para 243 em 2008.

5. POR QUE USAR AVALIAÇÃO POR MÉTODOS MISTOS?

Tradicionalmente, o argumento apresentado com mais frequência para justificar o uso de métodos mistos em avaliações é que a combinação de ferramentas quantitativas e qualitativas ajuda a mitigar as limitações do uso exclusivo de um ou de outro. Mais recentemente, o uso de métodos mistos tem sido visto como especialmente apropriado nos seguintes casos:

1. “reduzir o risco de vieses e contrabalançar as limitações do uso de um único método, aumentando assim a solidez das conclusões” (GERTLER et al, 2011, p. 119);
2. “gerar um entendimento mais completo, contextual, circunstancial e complexo do fenômeno de interesse, do que seria gerado por um estudo usando apenas um método” (GREENE, 2005, p. 410);
3. “aproveitar o melhor de cada um dos métodos, ao mesmo tempo em que são superadas suas deficiências” (DENZIN, 1989, p. 244);
4. “celebrar a igualdade das diferenças” (MCGEE, 2003, p. 135); e
5. “lançar relações e padrões interessantes, muitas vezes surpreendentes e algumas vezes inesperados, através de pesquisas qualitativas (além de perguntar ‘quanto?’) e estabelecer o grau de confiança nessas ‘hipóteses de trabalho’ com base na pesquisa quantitativa” (GABARINO & HOLLAND, 2009 p.11).

Como demonstrado pela ampliação da ideia de “métodos” e avançando na compreensão de por que os métodos mistos devem ser usados, as avaliações com métodos mistos representam uma inovação. Contribuem para a definição e a busca de novas abordagens caracterizadas por uma série de fatores, como estrutura epistemológica e conceitual ampliada, extensão, foco, análises de dados e público-alvo, conforme o quadro a seguir. >

VALOR AGREGADO DOS MÉTODOS MISTOS EM AVALIAÇÃO	
Campo	Valor agregado
Epistemologia	<ul style="list-style-type: none"> • Avança no entendimento e explica a complexidade dos programas com base no princípio de que os fenômenos são, normalmente, imprevisíveis e não explicáveis por mecanismos lineares de causas.
Estrutura conceitual	<ul style="list-style-type: none"> • Aumenta a triangulação, isto é, o uso das conclusões de diferentes fontes de dados, para responder à mesma pergunta da avaliação (quanto mais convergentes os resultados, mais sólidas as conclusões correspondentes). • Informa estratégias de coleta múltipla de dados e instrumentos que se alimentam mutuamente. Por exemplo: uma série de estudos de caso etnográficos sobre padrões de alimentação em casas de família, práticas de agricultura e custos permitiria a identificação de itens que poderiam ser incluídos em uma pesquisa subsequente com o objetivo de medir as mudanças no <i>status</i> nutricional como resultado de um programa de alimentação em escolas. • Facilita a validação cruzada entre diferentes estruturas conceituais e hipóteses sobre como ocorrem mudanças resultantes de intervenções. Por exemplo: como resultado da identificação (por meio de análise temática) de padrões de comportamento entre os beneficiários do programa, uma nova variável pode ser criada e incorporada em uma análise de regressão, aumentando assim o poder de previsão de um modelo estatístico subjacente ao programa em avaliação. • Oferece complementação aos resultados quantitativos e qualitativos. Por exemplo: alguns estudos de caso desenvolvidos a partir de um levantamento feito após um programa de treinamento poderiam explicar por que as mulheres participantes tinham quatro vezes menos chances de encontrar um trabalho nove meses depois de completar o programa que está sendo avaliado.
Extensão	<ul style="list-style-type: none"> • Combina o entendimento da extensão e profundidade das experiências humanas. • Aumenta a precisão dos resultados ao qualificá-los com base nos atributos temporais e espaciais de uma intervenção, além das características da(s) população(ões) afetada(s) pelo programa que está sendo avaliado.
Análise de dados	<ul style="list-style-type: none"> • Transformação de dados: dados numéricos são convertidos em narrativos e vice-versa, com o propósito de compreender e comunicar. • Análise de casos extremos: casos importantes de desempenho muito bom ou ruim são identificados e investigados mais a fundo a partir de um método diferente. • Desenvolvimento de teoria e teste: primeiro, identifica e analisa processos associados a indivíduos, comunidades e outras entidades classificadas de acordo com uma tipologia guiada pela teoria; segundo, inclui variáveis de categorias na análise estatística para ajudar a aumentar o poder de explicação do modelo desenvolvido. • Fusão de dados: variáveis qualitativas/categóricas e quantitativas são criadas e fundidas na mesma disposição de dados e usadas na análise estatística de um modelo explanatório.
Entendimento ampliado dos processos/ mecanismos que levam aos resultados	<ul style="list-style-type: none"> • Ajuda a abrir a chamada “caixa-preta”, combinando a identificação dos resultados finais de uma intervenção com aqueles da implementação de processos e dinâmicas (cultural, econômica, política, social) que influenciaram os efeitos dessa intervenção.
Amostragem	<ul style="list-style-type: none"> • Amostragens adicionais podem ajudar na formulação da questão da avaliação.
Público-alvo	<ul style="list-style-type: none"> • Flexibilidade na combinação e análise das conclusões garantem maior entendimento e aceitação dos resultados, aumentando a responsabilidade do trabalho de avaliação. • Quando as avaliações conduzidas de maneira participativa respondem a várias preocupações dos interessados (capturam nuances dentro das comunidades/indivíduos estudados) e interesses múltiplos.

Fonte: Tarsilla (2013).

A partir daqui serão exploradas diversas estruturas filosóficas que guiam a utilização de métodos mistos em avaliação, junto com exemplos ilustrativos e análises críticas dessas abordagens.

6. MÚLTIPLAS LENTES FILOSÓFICAS E TEÓRICAS NA AVALIAÇÃO

Alkin (2004) apresentou um retrato das raízes teóricas para a avaliação na América do Norte na primeira edição do *Evaluation Roots* (Raízes da Avaliação, em tradução livre), usando a metáfora de uma árvore para descrever os ramos das teorias de avaliação, isto é, “Métodos”, “Usos” e “Valores”. O ramo dos Métodos era povoado por aqueles que acreditavam que o rigor dos RCTs produziria conhecimento com credibilidade para os avaliadores.

Os teóricos do ramo do Uso perceberam que os avaliadores poderiam seguir rigidamente as regras baseadas no pós-positivismo, mas que o resultado de seus estudos não faria a menor diferença se ninguém os usasse. Por isso, seu foco era identificar os potenciais usuários e criar estudos que teriam credibilidade perante os mesmos.

Os teóricos do ramo dos Valores enfatizavam a importância do contexto e das construções de realidade dos diversos interessados como caminho para a criação de conhecimento que tivesse credibilidade. A primeira edição da *Evaluation Roots* foi descrita como algo que oferecia um bom começo do processo de documentar a evolução das teorias no campo da avaliação, mas foi criticada por não incluir avaliadores de fora da América do Norte e por não ter representações das perspectivas teóricas de profissionais de minorias étnicas ou indígenas (DAVIDSON, 2006; MERTENS, 2006).

Na segunda edição de sua obra, Alkin (2013) usou os mesmos ramos para descrever teóricos da avaliação, e dois capítulos adicionais foram acrescentados para incluir teóricos do resto do mundo. Primeiro, o capítulo de Stame (2013), sobre uma árvore da teoria da avaliação europeia que menciona a crescente necessidade de métodos mistos³ como ferramenta para encerrar a tensão entre um

3. O uso das revisões sistemáticas combinado às revisões narrativas defendidas por vários teóricos (PETTIGREW & ROBERTS, 2006) é um dos exemplos mais ilustrativos dos métodos mistos na teoria e prática de avaliação europeia.

paradigma “agrícola-botânico”, usando as tradições de testes em psicologia experimental e mental (PARLETT & HAMILTON, 1997), e o paradigma iluminista da avaliação, referenciado à antropologia, à psiquiatria e à sociologia⁴. Inspirado em Kushner (2000), Stame também lembra o leitor que é necessária uma mudança no pensamento avaliador de “documentar o programa e as vidas dos indivíduos nesse contexto” [para] “documentar as vidas e trabalho das pessoas e usar isso como o contexto no qual identificamos a importância e significado dos programas” (p. 359).

Da mesma forma, o capítulo de Rogers e Davidson sobre as raízes australianas e neozelandesas da avaliação menciona a relevância da combinação de métodos para capturar as diferenças substantivas das experiências humanas (frequentemente ditadas pela cultura de alguém, como no caso dos aborígenes e moradores das ilhas do estreito de Torres, passagem náutica entre o extremo norte da Austrália e a ilha de Nova Guiné) e tornando a avaliação mais sintonizada com a cultura na qual é aplicada (CRAM, 2009; DAVIDSON, 2010; KENNEDY & CRAM, 2010; WEHIPEIHANNA, DAVIDSON & MCKEGG, 2010).

Embora as contribuições dos autores da Europa, Austrália e Nova Zelândia tenham certamente ampliado a gama de teóricos da avaliação representados, ainda não haviam sido incluídas as vozes dos avaliadores de comunidades marginalizadas e de outros defensores de justiça social que vêm fazendo contribuições significativas no campo da avaliação. Assim sendo, Mertens (2012) e Mertens e Wilson (2012) acrescentaram um quarto ramo à árvore da avaliação identificado como Justiça Social. Mertens e Wilson também questionaram a metáfora da árvore com ramos separados e sugeriram uma metáfora baseada em correntes marítimas, que apresentam identidades distintas mas percorrem tra-

4. Uma abordagem de avaliação que parece apropriada para superar essa década de tensão é a Avaliação Realista. Ao reconhecer a possibilidade de identificar os mecanismos escondidos que influenciam os processos por trás dos processos (a caixa-preta) e ao reconhecer a existência de teorias “no meio do caminho” nos fenômenos mundiais, autores que adotam essas abordagens propõem uma alternativa tanto ao movimento positivista quanto ao construtivista na avaliação.

jetos comuns no fundo do mar, o que corresponderia ao compartilhamento das ideias.

Nós usamos esta estrutura de quatro ramos (sejam eles vistos como galhos de uma árvore, rios ou correntes marítimas) para estruturar nossa discussão sobre as bases para o uso de métodos mistos na avaliação. Nossa principal tese é que a compreensão dessas abordagens envolve a compreensão das posturas filosóficas, perspectivas teóricas e estratégias práticas que surgiram no mundo da avaliação associadas aos avanços nos métodos mistos.

Avaliadores passaram pelo que chamamos de “Guerra dos paradigmas”, quando os pós-positivistas discutiam com os construtivistas sobre que métodos resultariam na criação de evidências com credibilidade (MERTENS & WILSON, 2012). À primeira vista, as discussões pareciam ser sobre os métodos em si, mas eram, fundamentalmente, sobre as visões de mundo dos avaliadores e como enxergavam a axiologia (ética), a ontologia (realidade), a epistemologia (conhecimento) e a metodologia (pesquisa sistematizada) (SHADISH, 1998). Greene (2012, p. 756) identificou uma importante lição da “Guerra dos paradigmas”:

Uma importante lição do “grande debate quantitativo-qualitativo” foi o reconhecimento – em todas as divisões metodológicas – de que o conhecimento sobre assuntos humanos é parcialmente afetado pelas lentes e posturas do pesquisador humano. Ao aceitar a impossibilidade do empirismo puro e do conhecimento puramente objetivo sobre a ação humana proveniente do positivismo, os pesquisadores sociais tiveram que aceitar também a inevitável presença dele mesmo no conhecimento gerado pelas pesquisas sociais. Suas próprias crenças filosóficas e paradigmáticas sobre a realidade e o conhecimento – combinadas às suas teorias favoritas, às sensibilidades de seus mentores, às importantes experiências de vida, aos valores e crenças em geral – todos contam nas decisões metodológicas tomadas pelo pesquisador e na interpretação da trama do conhecimento obtido. (p. 756).

Com base na versão ampliada de Guba e Lincoln (1989; 2005) que descreve os quatro paradigmas (positivista, construtivista,

pragmático, teoria crítica), avaliadores operam atualmente em um mundo multiparadigmático que inclui: pós-positivistas, construtivistas, pragmáticos e transformadores (MERTENS, 2009; MERTENS, 2010; MERTENS & WILSON, 2012). Qualquer um desses paradigmas pode ser usado – individualmente ou combinados – como base para uma seleção de métodos mistos. Por isso usamos essa estrutura organizacional de paradigmas e ramos na próxima parte deste capítulo.

7. PARADIGMAS E RAMOS

Os quatro principais paradigmas se alinham aos quatro ramos da teoria da avaliação, como descritos na Tabela 6.1. Os pós-positivistas se alinham aos teóricos dos Métodos; os construtivistas com os teóricos do Valor; os transformadores com os teóricos da Justiça Social; e os pragmáticos com os teóricos do Uso. Os teóricos que escolhem trabalhar com vários paradigmas não aparecem na tabela a seguir, mas serão abordados adiante, quando chegarmos ao conceito de pluralismo dialético. >

PARADIGMAS FILOSÓFICOS E RAMOS DA AVALIAÇÃO	
Paradigma	Ramo
Pós-positivista	Métodos
Construtivista	Valores
Transformador	Justiça social
Pragmático	Uso

Fonte: adaptado de Mertens & Wilson, 2012.

7.1. O PARADIGMA PÓS-POSITIVISTA E O RAMO DOS MÉTODOS NA AVALIAÇÃO

A tabela a seguir mostra as crenças filosóficas associadas ao paradigma pós-positivista. Como suas crenças ontológicas e epistemológicas levam a uma metodologia que prioriza o uso de RCTs, perguntamo-nos como os métodos mistos entram nesse ramo. O recente trabalho de Howard White⁵ (2013) lida com essa questão

e esclarece que a pergunta-chave feita nas avaliações de impacto (“Que diferença fez a intervenção?”) pode ser mais bem respondida com RCT ou com formulações quase-experimentais se um RCT não é viável.

No entanto, ele também admite que as avaliações de impacto podem se beneficiar das abordagens qualitativas para obter respostas a diferentes tipos de questões (por exemplo: “Quais são as principais causas das mudanças observadas?”; “Qual a qualidade da implementação?”; “Quão relevante é a meta da intervenção que está sendo avaliada?”; “Quais as barreiras para a adoção de um novo comportamento ou da participação no programa em questão?”).

Ou, dito de outra maneira, as avaliações podem se fortalecer com o acréscimo de dados qualitativos (BELL, 2006; BRIGGS, 2006) a estudos essencialmente pós-positivistas (OAKLEY 2005a, 2005b). A perspectiva pós-positivista permanece como uma das mais utilizadas nas comunidades de pesquisa e avaliação da América do Norte e da Europa. Mas o fato de os métodos qualitativos serem considerados como “um suporte útil, ou segunda melhor opção”, atrás apenas do quantitativo (HESSE-BIBER, 2010, p. 457), não passou ao largo das críticas. Frequentemente repreendidos por não terem recebido treinamento avançado em metodologias qualitativas (HOWE, 2004), os pós-positivistas já foram acusados de introduzir visões e jargões excessivamente quantitativos na comunidade avaliadora, “marginalizando assim a pesquisa qualitativa de natureza aberta, com fluxo livre e emergente” (DENZIN, 2010, p. 420) e “deixando pouco espaço para questões ligadas ao empoderamento, à justiça social e a ‘políticas de esperança’⁶” (DENZIN, 2010, p. 420). >

5. Howard White é o diretor executivo da Iniciativa Internacional da Avaliação de Impacto (3ie), um programa especial que provê fundos e desenvolvimento de capacidade para a condução de avaliações de impacto em conjunto com estudos qualitativos e revisões sistemáticas das políticas de desenvolvimento, projetos e programas internacionais. Seu artigo “Será que vamos aprender?”, publicado pelo Centro de Desenvolvimento Global em 2007, enfatiza a relevância da avaliação de impacto e ressalta a natureza complementar dos métodos qualitativos para responder às questões da efetividade e impacto das intervenções de desenvolvimento internacional.

CRENÇAS DO PARADIGMA PÓS-POSITIVISTA	
O que se presume	Crenças
Axiologia	<ul style="list-style-type: none"> • Respeito/privacidade (confidencialidade) • Beneficência/não maleficência • Oportunidades iguais para participar e ser beneficiado, especialmente nos casos de RCT • Justiça (especialmente no caso de amostragem aleatória para evitar vieses e a eleição de participantes facilmente acessíveis)
Ontologia	<ul style="list-style-type: none"> • Uma realidade esperando ser descoberta e medida
Epistemologia	<ul style="list-style-type: none"> • Objetiva, neutra e distante
Metodologia	<ul style="list-style-type: none"> • Quantitativa; RCT ou formulações quase-experimentais (com o uso de técnicas estatísticas para compensar a falta de aleatoriedade); pesquisa estruturada

Fonte: adaptado de Mertens & Wilson, 2012.

O paradigma pós-positivista no mundo real: exemplos

Dada a quantidade de recursos alocados para rigorosas avaliações de impacto em uma variedade de setores na última década, sua prática contemporânea apresenta muitas posturas que se filiam ao paradigma pós-positivista e em abordagens fortemente baseadas em métodos. Visto como possível “aliança de interesses”, seja para a obtenção de financiamento adicional, seja para responder à preocupação do público sobre a aleatoriedade (isto é, fornecer bens e serviços a algumas comunidades, mas negá-los a outras em nome do rigor), a combinação de métodos quantitativos e qualitativos ganhou popularidade em diversos setores, especialmente no campo do desenvolvimento internacional, como ilustrado nos exemplos seguintes.

Em primeiro lugar, os métodos qualitativos vêm sendo cada vez mais usados nas avaliações de impacto conduzidas em todo

6. NT: Livre tradução da expressão “politics of hope”, as políticas de esperança são identificadas na plataforma do Partido Democrata Americano como aquelas voltadas a recuperar uma série de direitos aos cidadãos americanos, bem como a reposicionar o papel dos EUA na comunidade internacional. A expressão é marcadamente encontrada nos discursos do presidente Barack Obama, sendo inclusive título de um de seus livros, *The Audacity of Hope* (no Brasil, *A Audácia da Esperança*, pela Larousse Editora).

o mundo pelo Laboratório de Ação da Pobreza Abdul Latif Jameel (J-PAL, na sigla em inglês), afiliado ao Massachusetts Institute of Technology (MIT, Instituto de Tecnologia de Massachusetts), Estados Unidos. Ao longo de mais de cem RTC e avaliações de impacto conduzidas na última década, e com base na conclusão de que os resultados de avaliações unicamente quantitativas deixavam perguntas-chave avaliativas sem resposta, alguns economistas e quantitativistas da instituição começaram a usar (ainda que marginalmente) métodos qualitativos.

Tradicionalmente céticas sobre a utilidade de métodos não estatísticos para fins de pesquisa e avaliação, um grande número de avaliações de impacto produzidas pelo J-PAL em meados dos anos 2000 começou a usar métodos qualitativos, primeiramente como meio de identificar e explicar os processos por trás das causas dos impactos detectados em análises estatísticas (por exemplo, análises de correlação ou de regressão). Foi esse o caso na avaliação de impacto do projeto de desenvolvimento comunitário GoBifo⁷, financiado pelo Banco Mundial em Serra Leoa⁸ (KASEY et al, 2011). Ao longo dessa avaliação, a organização de grupos focais combinada à conversão dos resultados qualitativos correspondentes em variáveis e números – que serviram como base para o desenvolvimento das pesquisas seguintes – provou-se especialmente útil para explicar a dinâmica social dentro do contexto dos vilarejos atendidos pelo projeto.

Em segundo lugar, um exemplo ilustrativo dessa mudança de pensamento sobre os métodos dentro das agências, que favorecia os métodos quantitativos há décadas, é a contratação recente de um avaliador qualitativo no Gender Innovation Lab (GIL, na sigla em inglês, Laboratório de Inovação de Gênero), um novo escritório criado dentro da unidade de Gestão Econômica e de Redução de Pobreza (PREM, na sigla em inglês) da região africana do Banco Mundial. Com a missão de conduzir doze avaliações de impacto

7. GoBifo significa “Vá em Frente” na língua Krio.

8. O projeto oferecia fundos em bloco para comunidades para serem usados na construção de bens públicos como escolas, banheiros e locais para a secagem de grãos; e/ou financiamento ou treinamento de habilidades profissionais e atividades para geração de renda.

nas áreas de produtividade agrícola, empreendedorismo, emprego e direitos de propriedade entre 2012-2015⁹, o recém-contratado avaliador ajuda a equipe no desenvolvimento de estudos qualitativos e a integrar as questões relacionadas ao processo – e ao contexto – nas pesquisas quantitativas já existentes, com o propósito de medir de maneira mais eficiente o impacto das intervenções de desenvolvimento. Com base na conclusão de que, no passado, os estudos quantitativos não apontaram uma forma de dar às mulheres acesso consistente aos insumos econômicos, a equipe do GIL reconheceu o papel crítico que os métodos qualitativos desempenham na identificação de possíveis soluções¹⁰.

Em terceiro lugar, a variedade de avaliações predominantemente quantitativas conduzidas recentemente exemplificam alguns dos princípios-chave do paradigma pós-positivista: (a) a avaliação do impacto da reconstrução pós-conflito da Libéria, financiada pelo Departamento de Desenvolvimento Internacional do Reino Unido (DFID, na sigla em inglês) e Comitê Internacional de Resgate; (b) a avaliação de um programa de transferência de renda no Cazaquistão, financiada pelo Save the Children; e (c) a avaliação do impacto do trabalho de emergência e reabilitação da FAO (Organização para Alimentação e Agricultura da ONU) na zona rural da República Democrática do Congo (RDC).

Assim sendo, o paradigma pós-positivista e sua associação ao ramo dos Métodos proposto por Alkin (2004) exemplifica o uso de métodos mistos predominantemente quantitativos, em que o

9. Ao contrário das tentativas passadas para medir as desigualdades de gênero dentro de uma intervenção de desenvolvimento qualquer, essa nova geração de avaliação de impacto pretende identificar, desenvolver e testar soluções inovadoras para aliviar as limitações subjacentes de gênero.

10. No passado, foi por métodos qualitativos (por exemplo, estudos de caso e entrevistas aprofundadas e não estruturadas) que se identificou uma série de estratégias eficientes para combater as questões de gênero: (1) prover campos para que mulheres plantem os grãos que elas estão cultivando; (2) que tal área ficasse perto da casa das mulheres (para minimizar as limitações referentes a tempo e mobilidade); (3) a combinação do terreno com subsídios para insumo (devido às limitações de crédito disponível, reveladas nas entrevistas aprofundadas); e (4) o treinamento das mulheres mais influentes sobre como plantar cultivos de maior valor, tradicionalmente dominados pelos homens, sem alterar a extensão de terra disponível em seus programas.

componente qualitativo é usado para melhorar a compreensão do fenômeno estudado. Os exemplos refletem a crença de que uma medida quantitativa pode ser usada para capturar a “realidade” do impacto de um programa, e que os avaliadores, em sua maior parte, deveriam adotar postura neutra e objetiva. Na próxima seção, discutiremos as crenças do paradigma construtivista e como ele se alinha ao ramo dos Valores.

7.2. O PARADIGMA CONSTRUTIVISTA E O RAMO DOS VALORES DA AVALIAÇÃO

Como mencionado anteriormente, avaliadores construtivistas rejeitam a crença ontológica de que há uma realidade lá fora esperando ser medida em favor de uma postura que defende que as realidades são múltiplas e construídas socialmente (veja a tabela a seguir para a descrição das crenças filosóficas associadas ao paradigma construtivista). Essa crença enfatiza o uso de métodos qualitativos porque eles permitem ao avaliador interagir com os diversos interessados para juntos criarem diferentes versões da realidade social, necessárias para o total entendimento do contexto e do processo de implementação de uma intervenção. Ainda nos anos 1980, os construtivistas reconheceram que dados quantitativos poderiam ser incluídos como parte de um estudo qualitativo primário (GUBA & LINCOLN, 1989). Assim sendo, a porta tem estado e continua sendo aberta para considerar os estudos baseados em métodos mistos dentro do paradigma construtivista. >

CRENÇAS FILOSÓFICAS DO PARADIGMA CONSTRUTIVISTA	
O que se presume	Crenças
Axiologia	• Representação equilibrada; empatia; respeito multicultural
Ontologia	• Realidades múltiplas socialmente construídas
Epistemologia	• Interativa e interpretativa
Metodologia	• Abordagens primeiramente qualitativas

Fonte: adaptado de Mertens & Wilson, 2012.

Denzin (2012) sugere que os estudos com métodos mistos filiados ao paradigma construtivista têm maior potencial de produzir

achados que elucidem os objetos sociais do que aqueles baseados em métodos mistos mais próximos ao paradigma pós-positivista. Associados ao paradigma construtivista, os métodos mistos criam oportunidades de compreender os níveis contextuais e interpretativos da experiência nos quais os sentidos são produzidos, desenhando assim caminhos para promover justiça social. Os construtivistas veem a pesquisa sistemática como “um processo interativo delineado pela história pessoal, biografia, gênero, classe social, raça e enquadres étnicos das pessoas envolvidas” (p. 85). Denzin propõe uma moratória das “conversas sobre desenho e tipologia de métodos mistos para voltar à tarefa original, que é mudar o mundo” (p. 85). Sua posição é que pesquisadores precisam empregar estratégias construtivistas para criar textos que desafiem e estimulem os leitores à ação e defende fortemente o uso de paradigmas construtivistas como estrutura para a pesquisa engajada. No entanto, uma grande parte da comunidade construtivista não visualiza essa como a primeira meta de seu trabalho (MERRIAM, 2009). Muitos argumentam que o trabalho deveria ser descritivo e interpretativo, e não de ativismo. O paradigma transformador e o ramo da Justiça Social, que promovem os direitos sociais como sua principal premissa, serão discutidos na próxima seção.

Hesse-Biber (2013) nos traz mais material para pensar sobre como a lente construtivista, informada pela teoria crítica, pode ser usada em combinação com abordagens quantitativas, como nos RCT. Em contraste com a visão de White (2013) sobre o papel dos métodos qualitativos na avaliação, Hesse-Biber promove um papel mais forte para as investigações qualitativas nas avaliações com métodos mistos. Ela argumenta que a abordagem qualitativa pode ser usada mais extensivamente para responder perguntas como: Até que ponto a intervenção responde ao contexto e à cultura da população-alvo? Até que ponto funciona o trabalho de recrutamento? Até que ponto a população-alvo reflete a gama de objetivos finais do projeto? Quem ficou de fora? Por quê? Em que medida há práticas éticas (por exemplo: capacidade de responder às especificidades culturais, capacidade de reconhecer a desigualdade de poder) incluídas no processo de recrutamento? Em que medida a população-alvo entende aquilo que está consentindo?

Até que ponto os participantes aceitam os resultados da seleção aleatória? Estão dispostos a participar desses processos?

A autora argumenta que os RCT seriam melhorados se envelopados em uma perspectiva subjetivista, o que levanta uma questão fundamental: é possível que as avaliações com métodos mistos de inspiração qualitativa contribuam para que os pesquisadores tenham mais capacidade de demonstrar uma relação causal entre variáveis (MERTENS & HESSE-BIBER, 2012)?

O paradigma construtivista no mundo real: exemplos

Um crescente número de avaliações de impacto financiadas pelo DFID e outros doadores no norte da Europa (Agência de Desenvolvimento Internacional Sueca e Agência de Desenvolvimento Internacional Dinamarquesa) tem dado renovada proeminência ao paradigma construtivista nos últimos anos. A ideia de que métodos qualitativos ajudam a representar a pluralidade das perspectivas dos interessados em uma variedade de fenômenos foi certamente uma resposta às frustrações e limitações associadas aos esforços finitos dos RCT em medir ligações causais entre intervenções e efeitos. A literatura especializada parece ter sido especialmente útil na revitalização desse paradigma e na reafirmação de que métodos qualitativos têm de ser incluídos na caixa de ferramentas das análises causais (BRADY, 2002; PAWSON & TILLEY, 2007; YIN 2003).

Knigge e Cope (2006) ilustram a abordagem qualitativa em sua avaliação dos jardins comunitários na área mais empobrecida da cidade de Buffalo, estado de Nova York, Estados Unidos, para determinar sua contribuição para a acumulação de capital social. Fortemente apoiados por métodos qualitativos e tecnologias modernas, graças ao uso de uma plataforma virtual associada a sistemas de informação geográfica (GIS, na sigla em inglês), a equipe de avaliação criou primeiro um mapa mostrando os recursos disponíveis para a comunidade próximo dos jardins comunitários. Depois, uma representação visual dos jardins comunitários foi complementada com informações adicionais sobre a composição étnica dos moradores e os atributos do território, além de fotos da vizinhança e comentários das pessoas em áudio. Com base nas

especificações fornecidas pela população local, os responsáveis por planejar a avaliação concordaram em redesenhar seu mapa para melhor refletir a realidade revelada pela plataforma virtual. Os métodos incluíram a realização de entrevistas e observações, lançando mão de passeios de bicicleta na região-alvo como forma de medir a opinião das pessoas sobre o valor do projeto. O líder da equipe de avaliadores, Knigge, acrescentou estatísticas sobre o mercado de trabalho aos do GIS e usou a contagem de lotes disponíveis no mapa original para mostrar a associação dos jardins comunitários ao preço das casas. Como Fielding (2012) relatou em sua revisão sobre a integração de dados com novas tecnologias:

Knigge percebeu que ao considerar somente os dados quantitativos, ela “poderia ter ignorado a existência de jardins comunitários e todo um estudo etnográfico teria ignorado potenciais correlações e grupos que seriam melhor analisados pelo GIS” (KNIGGE & COPE, 2006, p. 2.934, apud FIELDING, 2012, p. 7).

Um segundo exemplo ilustrativo do paradigma construtivista em avaliações é a investigação do impacto do Programa de Reforma para o Desenvolvimento na Comunidade Gram Panchayat, na Índia (BAMBERGER, 2012). A partir da designação aleatória de 200 vilarejos (Gram Panchayats) para o projeto e de grupos de controle (estrutura conceitual quantitativa), a avaliação baseou-se no estudo exploratório da posse da terra, propriedade dos bens públicos, participação e redes sociais (coleta de dados qualitativos), que serviram para criar a base da pesquisa realizada antes do início da série de treinamentos previstos no programa. A essas primeiras e poucas atividades seguiu-se a coleta de dados qualitativos em cinco projetos e cinco áreas de controle para descrever os processos de mudanças de comportamento (por exemplo: novas estratégias de organização das comunidades, tipos e variedades de projetos de desenvolvimento comunitários, extensão da participação das mulheres e participação das castas). A avaliação incluiu visitas de campo e entrevistas nas comunidades, permitindo a descoberta dos processos pelos quais as dinâmicas política e social, a corrupção, a mudança econômica e a afiliação a redes impactaram a

efetividade da iniciativa, medidos na comparação entre os efeitos antes e dois anos depois de sua implementação.

Um terceiro exemplo de construtivismo é uma avaliação com métodos mistos conduzida em um estado majoritariamente rural na Índia, realizada para examinar as experiências das mulheres com indução de abortos, seus determinantes e sua prevalência. A abordagem combinou pesquisa com métodos quantitativos, que incluía perguntas abertas, grupos focais e estudos de caso voltados a compreender as tentativas de aborto e os abortos realizados, investigando tanto as motivações pessoais quanto fatores culturais associados aos casos. Edmeades escreveu o seguinte:

A abordagem utilizada diminuiu significativamente o não relato de abortos, seja porque ofereceu às mulheres um espaço seguro para discutir a questão, seja porque melhorou a memória delas em relação a eventos do passado. Como resultado, os dados sobre abortos e as circunstâncias em que foram realizados são, provavelmente, mais representativos da totalidade das experiências com interrupção de gravidez do que os dados coletados em abordagens mais tradicionais. Isso tem o efeito de aumentar a validade e confiabilidade das conclusões alcançadas por meio da análise de dados, que são de particular importância para pesquisadores e responsáveis por políticas públicas (EDMEADES et al, 2010, p. 194).

O paradigma construtivista e sua representação no ramo dos Valores da avaliação são demonstrados por uma amostra de estudos por métodos mistos, cujas formulações foram não apenas qualitativamente dominantes, mas focaram-se na construção do entendimento da realidade dos participantes por meios que permitiam que viesse à tona a complexidade de suas experiências. Dados quantitativos também contribuíram para a qualidade dos estudos, mas eles foram apurados a partir do conhecimento gerado pelos métodos qualitativos. A próxima seção discute o paradigma pragmático e o ramo do Uso da avaliação. O foco se dirige aos principais usuários das avaliações e à produção de estudos capazes de atender às suas necessidades.

7.3. O PARADIGMA PRAGMÁTICO E O RAMO DO USO DA AVALIAÇÃO

O paradigma pragmático (MORGAN, 2007; JOHNSON & ONWUEGBUZIE, 2004) defende o uso de métodos mistos baseado na crença de que não há apenas um grupo de métodos apropriados; ao contrário, o critério para a escolha dos métodos inclui identificar aqueles que melhor se adequam às questões da avaliação. As crenças filosóficas associadas ao paradigma pragmático, como são às vezes descritas na literatura dos métodos mistos, aparecem na tabela a seguir. >

CRENÇAS FILOSÓFICAS DO PARADIGMA PRAGMÁTICO	
O que se presume	Crenças
Axiologia	<ul style="list-style-type: none"> Alcançar conhecimento na busca dos fins desejados e melhorias na sociedade conforme influenciado pelos valores, políticas e experiência dos avaliadores
Ontologia	<ul style="list-style-type: none"> A realidade é continuamente criada com a experiência de interação e transação com o “mundo”
Epistemologia	<ul style="list-style-type: none"> Ideias e conhecimentos são avaliados de acordo com suas consequências; o melhor resultado que os avaliadores podem conseguir é a garantia de assertividade e verdades provisórias sobre os objetos
Metodologia	<ul style="list-style-type: none"> Combina os métodos às questões; métodos mistos

Fonte: adaptado de Mertens & Wilson, 2013.

Biesta (2010), Greene e Hall (2010), Hall (2013), e Denzin (2012) advertem contra uma aplicação excessivamente simplista do pragmatismo como filosofia de avaliação, como a que expressa na ideia “se o método se encaixa na questão, use-o”. Biesta (2010) expõe os princípios básicos do pragmatismo de Dewey como uma filosofia capaz de orientar avaliações com métodos mistos, à medida que Dewey defendia que nenhum processo de conhecimento pode ser considerado suficientemente sólido para se constituir como “verdade”. Ao contrário, diferentes formas de se engajar com o mundo social geram diferentes padrões de conhecimento.

Denzin (2012) pergunta: “Será que os defensores dos métodos mistos cometeram uma injustiça contra o pragmatismo como enquadre filosófico para os métodos mistos?”. Hall (2013) oferece uma visão com mais nuances do paradigma pragmático com base no trabalho de Dewey, ampliando a compreensão. Ela defende o

uso do pragmatismo de Dewey porque este enfatiza a reflexão, a ética, a justiça social e o uso de pesquisa sistemática a serviço da solução de problemas da sociedade, com a adoção de ações inteligentes. Para poder implementar tais ações, os avaliadores precisam interagir com as comunidades onde trabalham e estar abertos a reflexões críticas sobre os significados gerados em termos de identificação de problemas e potenciais soluções. Essas não são vistas como entidades estagnadas; ao contrário, os avaliadores têm a responsabilidade de coletar dados sobre as consequências das soluções de forma contínua e de interagir com membros da comunidade da mesma forma, para garantir que sejam feitas as adaptações necessárias a aumentar o sucesso da empreitada. Assim sendo, a coleta de evidências tem de ser contínua, responder às demandas e ser multifacetada. Essa seria a razão fundamental para o uso de métodos quantitativos e qualitativos em avaliações pragmáticas. Dewey enfatiza ainda a importância das consequências das intervenções na vida das pessoas num sentido democrático, concluindo que programas sociais devem melhorar a vida de todos os interessados, tanto os privilegiados quanto os marginalizados.

O paradigma pragmático: exemplos do mundo real

Robinson et al (2011) apresentam um exemplo de estudo com métodos mistos que se alinha às crenças do paradigma pragmático. O estudo foi formulado para avaliar as necessidades e experiências de famílias que cuidavam de seus membros com demência. A justificativa da escolha dos métodos foi a seguinte:

Julgamos que nossa medição objetiva dos pacientes com demência e de seus cuidadores só teria significado se fosse levado em conta o contexto das famílias. Isso exemplifica por que uma abordagem com métodos mistos se encaixa num estudo sobre demência: problemas complexos pedem abordagens de investigação também complexas. A demência é uma condição profundamente complicada e o uso de métodos mistos permite uma gama diversa de meios para se lidar com o problema (p. 311).

Eles mencionam especificamente ter escolhido o paradigma pragmático porque, de acordo com Morgan (2007), ele oferece uma base para o uso de métodos quantitativos e qualitativos sem precisar dar atenção a preocupações metafísicas como a ontologia e a epistemologia. Também descrevem o paradigma pragmático como algo que lhes permite mover-se entre componentes dedutivos, quantitativos, e indutivos, qualitativos, do estudo.

Um segundo exemplo do paradigma pragmático é a avaliação feita em múltiplos níveis com múltiplos métodos do programa de microempréstimos administrado por uma agência de financiamento canadense, cujo alvo principal eram indivíduos desempregados interessados em criar ou ampliar seus próprios negócios (JACKSON & TARSILLA, 2012, TARSILLA 2010). Vários métodos foram usados para responder tanto a questões sobre a efetividade do programa (“Até que ponto os beneficiados fortaleceram seus negócios como consequência de sua participação no programa?”; “Até que ponto os microempréstimos concedidos contribuíram para o bem-estar de suas famílias?”) e sobre processos de implementação (“Que aspectos da relação entre banqueiros e beneficiados poderiam ser fortalecidos no futuro próximo?”; “Como o governo local poderia ajudar iniciativas similares no futuro?”). A coleta de dados compreendeu três níveis: a) micro: os indivíduos que pegaram os microempréstimos e suas famílias; b) médio: os funcionários da financiadora; e c) macro: os funcionários dos governos local e federal. Os métodos de coleta de dados foram:

- Entrevistas com funcionários-chave que distribuíam os microempréstimos na agência financiadora e com professores da Universidade de Carleton com *expertise* em desenvolvimento econômico de comunidades (CED, na sigla em inglês) e retorno social do investimento (SROI, na sigla em inglês);
- Revisão de documentos sobre microfinanças e CED no Canadá e no mundo (inclusive auditorias e relatórios anuais conduzidos no passado);
- Questionário semiestruturado e grupos focais: usados para abordar perguntas relacionadas a indicadores-chave de resultados (emprego, situação de crédito, assistência social e de saúde etc.);

- Questionário on-line semiestruturado com perguntas abertas e fechadas;
- Grupo de discussão virtual para validar conclusões preliminares;
- Dez estudos de caso com base em entrevistas individuais realizadas com sete tomadores de empréstimo inadimplentes e três inadimplentes.
- Cálculo da Declaração de Valor Agregado Expandido (EVAS, na sigla em inglês) para quantificar o valor “oculto” das contribuições não monetárias do programa (MOOK, QUARTER & RICHMOND, 2007), tais como a oferta gratuita de sessões de educação financeira e as matérias publicadas pela mídia local.

Um terceiro exemplo foi a avaliação do processo de socialização de crianças e de seus padrões de circulação no território de um município na Dinamarca. Como parte desse estudo, a equipe combinou trabalho de campo etnográfico com a tecnologia GPS (sistema de posicionamento global) usando um questionário interativo que as crianças respondiam via celular (CHRISTENSEN et al, 2011). A metodologia inovadora usada nessa avaliação procurou enfatizar as experiências subjetivas das crianças com observações sistemáticas, mapeamento e aplicação de questionários. Nas palavras da equipe responsável:

Nosso grupo se formou a partir de um reconhecimento quase coincidente: dois pesquisadores do campo da mobilidade urbana sentiram necessidade de incluir contribuições qualitativas da ciência social para entender a base subjetiva e contextual do movimento das pessoas pela cidade; e dois antropólogos precisavam entender o padrão diário da mobilidade infantil. Portanto, neste estudo, uma abordagem pragmática, que acentua a flexibilidade dos métodos e técnicas, aliada à colaboração interdisciplinar levou à formulação de uma pesquisa integrada que uniu nossos esforços em uma abordagem holística (CHRISTENSEN, 2011, p.230).

O último exemplo relacionado ao paradigma pragmático é a avaliação de impacto dos processos comunitários integrados de tomada de decisão conduzidos pelo National Democratic Institute

(NDI, na sigla em inglês). A coleta de dados para pesquisa combinou entrevistas em profundidade com membros da comunidade aleatoriamente escolhidos, com o objetivo de aferir mudanças de conhecimento, opinião e atitudes de funcionários públicos locais a respeito de mecanismos de governança. Os métodos qualitativos foram utilizados nessa rigorosa avaliação de impacto para identificar qualquer aspecto inesperado e/ou fatores externos que pudessem prejudicar a efetividade do projeto. Os resultados foram bastante interessantes: enquanto as respostas aos questionários sugeriam que o programa estava alcançando os objetivos esperados, as entrevistas em profundidade e os grupos focais evidenciaram que funcionários públicos corruptos e a apreensão injustificada de terras das comunidades prejudicaram gravemente o sucesso do projeto, questionando o valor democrático dos processos de tomada de decisão que haviam sido implementados.

Os quatro exemplos anteriores ilustram as crenças do paradigma pragmático e do ramo dos Usos em avaliações onde o objetivo primordial é melhorar a utilização dos resultados pelos principais interessados. Embora algumas investigações envolvam membros dos públicos-alvo, a ênfase está no uso das informações pelos atores que detêm poder de decisão. O paradigma transformador e o ramo da Justiça Social, tratados a seguir, posicionam o uso das avaliações para melhorar a vida dos membros marginalizados da sociedade.

7.4. O PARADIGMA TRANSFORMADOR E O RAMO DA JUSTIÇA SOCIAL DA AVALIAÇÃO

As crenças filosóficas do paradigma transformador (MERTENS, 2009; MERTENS & WILSON, 2012) emanam de uma postura ética que enfatiza a busca por justiça social e, em última instância, a garantia de direitos humanos¹¹. Espera-se aqui que os avaliadores, a partir de suas crenças éticas fundamentais, elejam critérios que

11. Sweetman et al (2010) realizaram uma revisão da literatura ou de estudos a respeito de métodos mistos conduzidos a partir de lentes transformadoras. Foi identificado um grande número (272) deles. A forma mais popular de integração de métodos foi a sequencial, em que um grupo de dados se estendia, ou acrescentava, a outro grupo de dados.

impliquem certa compreensão sobre a realidade, sobre o papel do conhecimento e sobre o sentido de investigar a realidade de forma sistemática. Nesse sentido, a natureza da realidade é tomada como multifacetada e como reflexo das diferentes posições de poder dos sujeitos na sociedade. Um resumo das crenças filosóficas transformadoras é apresentado na tabela a seguir. >

CRENÇAS DO PARADIGMA TRANSFORMADOR	
O que se presume	Crenças
Axiologia	<ul style="list-style-type: none"> • Respeito cultural; promoção de justiça social e dos direitos humanos; combate às desigualdades; reciprocidade
Ontologia	<ul style="list-style-type: none"> • Multifacetada; consequências do privilégio
Epistemologia	<ul style="list-style-type: none"> • Interativa; confiança; ponto de vista
Metodologia	<ul style="list-style-type: none"> • Transformadora; por diálogo; emancipatória; métodos mistos

Fonte: adaptado de Mertens & Wilson, 2012.

As crenças axiológicas do paradigma transformador priorizam a igualdade de direitos e a justiça social, levando os avaliadores a fazer perguntas como: “Se eu iniciar minha avaliação com o objetivo de garantir direitos humanos e promover justiça social, o que isso significa sobre como eu vejo a realidade, o papel do conhecimento e o sentido de investigar a realidade de forma sistemática?”; e “Como essa postura deverá influenciar minhas crenças filosóficas e minha prática como avaliador?”. Uma preocupação imediata que surge dessa primeira postura ética é a necessidade de entender a cultura da comunidade em toda sua complexidade, inclusive as diferenças de poder associadas às posições sociais, culturais e econômicas. Outras questões são a importância de enxergar os ativos de uma comunidade e propiciar condições para que os membros de uma equipe de avaliação sejam encorajados a trazer seus pontos fortes para o estudo. Além disso, os estudos devem ser formulados de forma a permitir a reciprocidade: “O que os membros da comunidade ganham com a avaliação?”.

As crenças axiológicas transformadoras levam à consideração das crenças ontológicas e epistemológicas associadas a este paradigma. Os avaliadores precisam se engajar na coleta de dados de maneira a respeitar a cultura, permitindo que múltiplas versões

da realidade venham à tona (ontologia) e que sejam construídas relações de confiança (epistemologia). Assim sendo, os métodos mistos não são um requisito para a realização de avaliações transformadoras, mas eles costumam gerar melhores oportunidades de compreensão das comunidades em sua complexidade quando comparados ao uso de apenas um método.

No paradigma transformador, a realidade é vista como construída com base em diferentes posições sociais e culturais, como as de gênero, raça, etnia, pobreza, deficiência, surdez, religião, idade, indigência, situação migratória e identidade sexual. O avaliador tem então responsabilidade de revelar tais versões, expondo suas fontes e as consequências de se privilegiar uma versão em detrimento de outra. As diferentes versões surgem de diferentes situações de acordo com privilégios conquistados e não conquistados, assentados nas posições sociais e culturais antes mencionadas. Por exemplo: pessoas sem deficiência auditiva podem ter uma versão da realidade de que ouvir é melhor do que não ouvir. Esse entendimento contrasta com o entendimento dos deficientes auditivos, para os quais ser surdo significa ser membro de um grupo cultural linguístico minoritário. Se for privilegiada a versão de realidade das pessoas sem deficiência auditiva, isso pode resultar em uma negação dos direitos das pessoas com deficiência auditiva em termos de uso de linguagem visual, como a língua americana ou britânica de sinais. Foi exatamente isso que aconteceu nos anos de 1800, no Congresso Internacional de Educação de Surdos (Iced, na sigla em inglês), quando as pessoas sem deficiência auditiva decidiram que uma abordagem oral para a comunicação, que envolvia leitura labial e fala, era o melhor método para educar deficientes auditivos. A consequência da aceitação da visão de realidade dos não surdos é que, por mais de um século, deficientes auditivos em todo o mundo tiveram negado o acesso à linguagem de sinais em sua Educação. Resta alguma dúvida sobre por que os alunos surdos que se formavam nas escolas americanas tinham níveis primários de leitura?

Os avaliadores precisam conhecer todas as versões de realidade associadas aos privilégios não conquistados dentro das comunidades em que trabalham e precisam tornar visíveis as versões que mantêm um *status quo* opressivo e aquelas que aumentam a

possibilidade de estender o alcance dos direitos humanos. Métodos mistos proveem mecanismos para revelar essas diferentes versões, suas fontes e as consequências de se privilegiar uma versão sobre a outra. Por exemplo: uma avaliação sobre o acesso ao sistema judiciário para pessoas surdas e com deficiência auditiva nos Estados Unidos usou coleta de dados qualitativos em grupos focais com pessoas surdas para revelar a gama de experiências que eles tiveram nos tribunais (MERTENS, 2009). Suas experiências foram influenciadas pelo grau de deficiência auditiva, uso da linguagem (por exemplo: língua americana de sinais, língua mexicana de sinais, inglês falado) e sua habilidade de se beneficiar de tecnologias de apoio (por exemplo: aparelhos auditivos, implantes cocleares, legendagem em tempo real). Essa diversidade também precisou ser refletida no treinamento dos funcionários de tribunais e na avaliação quantitativa da implementação de planos para aumentar a acessibilidade nos tribunais.

Epistemologicamente, avaliadores transformadores estão cientes das relações de poder nas comunidades em que trabalham e da necessidade de estarem em constante estado de aprendizagem sobre como interagir de maneira culturalmente apropriada. A natureza do conhecimento dentro do paradigma transformador inclui as diferentes maneiras de conhecer inerentes a cada grupo interessado. Por exemplo: formas de conhecimento indígena foram exploradas por Chilisa (2012), Cram (2009) e outros. O processo envolve o estudo de uma história de opressão, tratados violados, roubo de terras e colonialismo que pode não parecer relevante para um avaliador não indígena. Mas esses saberes vêm sendo transmitidos por membros de comunidades indígenas e podem ofuscar questões que os responsáveis por desenvolver e avaliar programas tenham em mente. Métodos mistos podem ser importantes para a obtenção de bons dados, tanto pela possibilidade de responderem a rituais ou práticas culturais dessas populações, quanto por lançarem mão de instrumentos linguisticamente adequados para a coleta de dados.

Metodologicamente, recomenda-se uma formulação cíclica para a investigação, de modo a garantir o entendimento da complexidade cultural de uma comunidade. É preciso desenhar meca-

nismos para o trabalho em conjunto, registrar o conhecimento já existente e desenhar processos de coleta de dados que sejam vistos como legítimos e esperançosos para os membros da comunidade. Isso pode implicar o início de uma investigação qualitativa por meio de diálogos entre os grupos interessados. Pode também permitir a coleta desagregada de dados demográficos ou de incidência de diferentes tipos de variáveis relevantes sobre diversidade (por exemplo: gênero, filiação tribal, idade, língua). Esses dados qualitativos e quantitativos devem ser mostrados à comunidade para guiar a interpretação e formular os próximos passos.

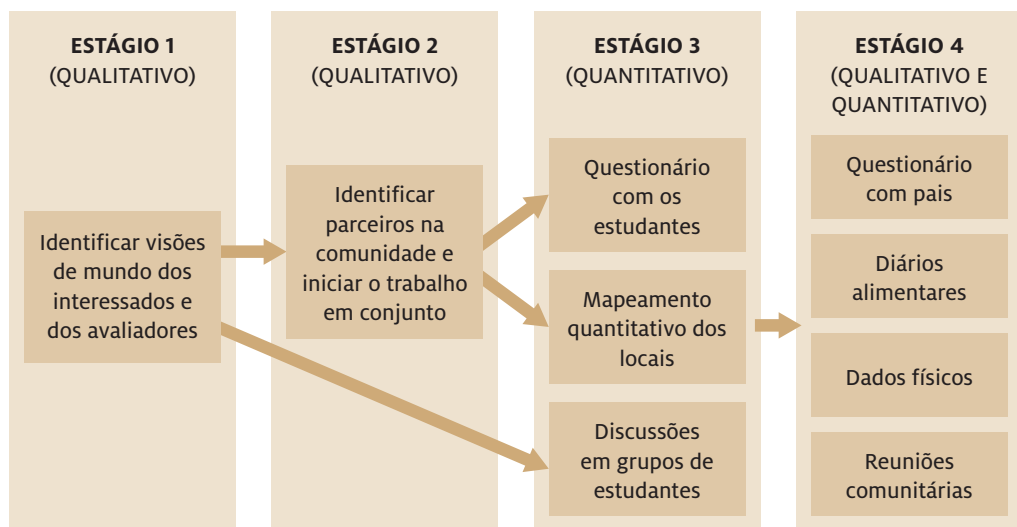
O paradigma transformador no mundo real: exemplos

Bledsoe (2010) conduziu um estudo transformador, usando métodos mistos, sobre um programa de redução de obesidade desenvolvido em uma cidade com alto índice de pobreza e muitos membros de minorias étnicas e raciais, além de imigrantes da África e do Leste Europeu. Quando ela se juntou aos gestores, disseram-lhe que os motivos para a obesidade nas escolas era a autoestima dos estudantes, reconhecida como muito baixa. Formulou-se então um programa para elevar a autoestima dos jovens e, quando Bledsoe perguntou se havia dados para comprovar essa teoria, a resposta foi negativa. Como primeiro passo, ela usou uma combinação de métodos qualitativos e quantitativos para averiguar as características demográficas dos estudantes e suas perspectivas sobre obesidade.

Os resultados revelaram que eles se sentiam bem sendo obesos, mas queriam evitar problemas de saúde associados à obesidade, como doenças cardíacas ou diabetes. Portanto, foi necessário formular um programa diferente do que o inicialmente planejado. O diagrama a seguir ilustra o desenho cíclico transformador criado com métodos mistos. Ele mostra que os avaliadores enfatizaram a importância de entender a cultura da comunidade e as forças ali atuantes que dificultavam escolhas alimentares saudáveis. O desenho é cíclico porque as informações reveladas a cada etapa foram usadas como base para a próxima. Por exemplo: uma vez entendido que certas barreiras estavam prejudicando a possibilidade de os estudantes perderem peso, um mapeamento quantitativo foi usado

para apontar os locais de maior dificuldade de acesso a frutas frescas e legumes, bem como a inexistência de lugares seguros para a prática de exercícios. As intervenções incluíram negociar com os comerciantes do bairro maior oferta de alimentos saudáveis; o uso de pedômetros pelos jovens durante uma dança para contar o número de passos realizados; e festivais culinários que ensinavam como comidas tradicionais podiam ser preparadas de maneira mais saudável. >

OS CICLOS DA AVALIAÇÃO TRANSFORMADORA DE UM PROGRAMA VOLTADO À REDUÇÃO DA OBESIDADE



Outro exemplo de avaliação transformadora com métodos mistos é o trabalho conduzido por Chilisa (2005; 2012) em Botswana. Seleccionada para avaliar o impacto de uma intervenção voltada a reduzir o contágio por HIV/Aids na África, ela começou questionando a formulação do projeto e, mediante coleta de dados e análise, conseguiu demonstrar uma variedade de mal-entendidos sobre a cultura e a língua do país. Como provado tanto pelos questionários quanto pelas entrevistas em profundidade com jovens, Chilisa demonstrou que as dinâmicas de poder entre homens e mulheres ocorriam de tal forma que elas jamais conseguiriam pedir a eles que usassem preservativos. Da mesma maneira, ela

conseguiu enfatizar que o uso de campanhas de orientação com mensagens em inglês não tinha efetividade, já que aqueles com maior risco de transmissão do HIV só falavam as línguas locais.

Outros exemplos interessantes de avaliações transformadoras com métodos mistos são fornecidos pelo crescente número de estudos sobre igualdade, financiados por várias agências da Organização das Nações Unidas (ONU) nos últimos anos. No caso da avaliação de resultados do programa Tostan do Unicef (que tem o objetivo de diminuir a circuncisão feminina em vilarejos do Senegal), por exemplo, os métodos mistos provaram-se cruciais para determinar o impacto total da intervenção. Em particular, o uso quantitativo de uma pesquisa nas casas foi combinado a observações, grupos focais e entrevistas em profundidade, com o objetivo de avaliar o processo de implementação, as dinâmicas em torno do compromisso das comunidades de abandonar a prática da circuncisão e a percepção do impacto do programa entre as mulheres que eram alvo da iniciativa.

Outro exemplo similar é a Avaliação em Tempo Real da resposta humanitária à crise de deslocamento do Paquistão em 2009 (BAMBERGER & SEGONE, 2012). Graças ao uso de métodos mistos, as conclusões apontaram para o fato de que a iniciativa havia sido planejada consultando apenas os anciões dos vilarejos e os chefes de família masculinos (eles formavam a maioria dos que responderam à pesquisa e participaram dos grupos focais), e que não foi dada nenhuma atenção às necessidades especiais das mulheres e crianças das famílias mais pobres e vulneráveis.

A avaliação do programa de Educação do Unicef para o Timor Leste (2003-2009) também ilustra bem o uso transformador de métodos mistos e confirma como a combinação de métodos quantitativos e qualitativos, além de servir ao propósito da igualdade, foi chave para compensar o acesso limitado aos dados quantitativos do país. Um último exemplo transformador é a avaliação do impacto da assistência social para diminuir a pobreza infantil e a exclusão social de crianças na Albânia. Ao longo desse processo, o garimpo de dados nacionais permitiu a identificação de diferentes grupos vulneráveis, que foram então entrevistados e produziram recomendações capazes de apoiar os gestores a melhorar diferentes mecanismos da rede nacional de proteção social.

8. COMBINANDO PARADIGMAS: O PLURALISMO DIALÉTICO

Uma abordagem dialética estabelece um diálogo entre suposições, práticas, entendimentos, pontos de vista e conclusões de um ou mais estudos que emanam de diferentes posições paradigmáticas (GREENE & HALL, 2010; JOHNSON, 2011). Greene (2012) explica os complexos desafios que os avaliadores encontram em termos de contexto, valores e papel ao investigar na sociedade. Ela destaca que diferentes tradições metodológicas na comunidade avaliativa respondem diferentemente a esses desafios e, por sua vez, moldam “o conhecimento gerado por um estudo e os fundamentos deste conhecimento... [Ela] apresenta uma resposta metodológica mista para esses desafios, sob o argumento primordial de que tal abordagem oferece oportunidades dialógicas para gerar melhor entendimento de fenômenos sociais importantes, precisamente porque legítima e respeita as múltiplas respostas a essas questões críticas e convida ao diálogo entre elas” (p. 755).

O pluralismo dialético (DP, na sigla em inglês) é uma postura filosófica que permite aos avaliadores engajarem-se em avaliações por métodos mistos enquanto vivenciam as tensões que surgem quando diferentes posturas paradigmáticas passam a dialogar. Johnson e Stefurak (2013) descrevem o DP como capaz de incluir uma ontologia pluralista confiando nos processos dialético, dialógico e hermenêutico que caracterizam a contínua interação entre atores com interesses diferentes (por exemplo: na filosofia, paradigma, cultura, metodologia, método). Os métodos dialéticos engajam atores que aparentemente detêm posições díspares em um diálogo significativo, usando múltiplas estratégias sociais, psicológicas, dialógicas e de negociação. Eles identificam as seguintes características no paradigma: construção de equipes heterogêneas e inclusivas; equalização de poderes; construção conjunta de critérios e parâmetros de julgamento; e compromisso com a democracia deliberativa e transformadora. O DP pode incluir observação, experiência, experimentos etc., com a expectativa de que dados empíricos múltiplos *não* vão convergir tipicamente em uma conclusão simples, mas que frequentemente fornecerão divergências que resultarão em melhor compreensão do objeto. A maioria dos fenômenos humanos é complexa e portanto existem muitas

perspectivas, pontos de vista e verdades parciais em torno de quase todos os objetos de estudo.

O paradigma do pluralismo dialético no mundo real: exemplos

Um exemplo ilustrativo de um estudo de DP com métodos mistos é a avaliação do Projeto de Demonstração de Gestão de Pessoal do Departamento de Agricultura dos Estados Unidos (USDA, na sigla em inglês) (MARK et al, 1997). O propósito da avaliação era testar a efetividade da nova estratégia de gestão de recursos humanos (RH) com o objetivo de melhorar os processos de recrutamento e seleção para a contratação de funcionários federais (por exemplo: a descentralização e simplificação das tarefas administrativas e responsabilidades associadas ao processo de contratação em duas áreas da USDA).

Primeiramente, com base nos dados quantitativos (isto é, revisão de informações sobre ações dos funcionários e candidatos e sobre os indivíduos contratados para cada vaga), a avaliação fez uso extensivo de métodos qualitativos (uma teoria de mudança também foi desenvolvida antes do início do trabalho, baseada em entrevistas semiestruturadas, revisão de documentos e um total de 95 visitas subsequentes a 45 locais). Várias características de pluralismo dialético foram identificadas no caso. Primeiro, a divergência das conclusões entre as respostas à pesquisa, que indicavam que os gestores apreciavam o valor agregado do novo sistema de contratação, e às entrevistas pessoais, demonstrando que eles não estavam satisfeitos com o novo sistema. A pluralidade de vozes discordantes entre os usuários finais da nova estratégia de RH levou à revisão do instrumento e à substituição da expressão “sistema de contratação” (que gerava confusão e mal-entendidos) por um rótulo mais preciso. Em segundo lugar, o sistema de entrevistas empregado pela USDA permitia aos avaliadores qualificar até que ponto a relação entre o candidato e a pessoa de RH influenciou o nível de satisfação do funcionário com a experiência de contratação como um todo (a análise de dados qualitativos confirmou a relevância dessa relação, enquanto a análise de dados quantitativos levou erroneamente a crer que se tratava de um “erro de variância”).

Outro exemplo de DP com métodos mistos é a avaliação da ideologia de masculinidade conduzida entre vários grupos étnicos na África do Sul (LUYT, 2011). Como parte dessa avaliação, um instrumento quantitativo (o MANI-II, uma medição multidimensional da ideologia de masculinidade) desenvolvido em 2001 foi revisto com base em resultados qualitativos que reconheceram diferenças conceituais básicas entre os grupos culturais no entendimento da masculinidade. Baseado em uma análise temática das transcrições de grupos focais, os resultados qualitativos do estudo apontaram falta de equivalência de conteúdo entre as culturas estudadas nas três versões da ferramenta (em inglês, afrikaans e xhosa¹²). Simultaneamente, foi conduzida uma análise fatorial do instrumento para aferir a validade de sua construção cultural. Ao fim da fase de coleta de dados quantitativos, analisou-se grande número de respostas (434 em afrikaans, 890 em inglês e 273 em xhosa) e também foi calculado o coeficiente de congruência (Pi de Tucker). Como escreveu o autor:

[...] dados qualitativos “descritivos de um pequeno grupo” e dados quantitativos “normativos de um grande grupo” facilitaram a rigorosa exploração da perspectiva de vários participantes entre e dentro dos grupos culturais [...] Dados quantitativos foram considerados antes dos dados qualitativos. Ainda assim, os dados qualitativos não foram considerados simplesmente suplementares, já que eles proporcionavam informação única, além de suplementar. Isso contribuiu para a noção de que formulações de status igual são possíveis em termos da soma de suas contribuições individuais. (EDMEADES, NYBLADE et al, 2010, p. 190).

Em resumo, o DP depende tipicamente da inclusão de métodos que provavelmente gerarão diferentes entendimentos e perspectivas dos avaliados. A construção proposital de uma equipe participativa de interessados múltiplos divergentes (incluindo aqueles com menos poder) é uma estratégia-chave para incluir as diferenças das vozes na avaliação.

12. NT: três das diversas línguas oficiais faladas na África do Sul.

9. QUESTÕES FUTURAS NA AVALIAÇÃO COM MÉTODOS MISTOS

Talvez porque as avaliações estejam inerentemente situadas “no mundo”, os profissionais do campo abraçaram o uso dos métodos mistos antes que fossem desenvolvidas estruturas filosóficas ou teóricas que tratassem explicitamente do significado de seu uso em vários níveis. Atualmente, mais do que aplicar métodos mistos como um padrão, os avaliadores têm à disposição uma rica gama de opções – do nível filosófico ao nível da prática – para estimular o pensamento crítico necessário a estudos complexos.

As diferenças de opinião sobre as metodologias apropriadas e a credibilidade das evidências têm lastro nas premissas filosóficas dos profissionais, como foi demonstrado neste artigo. E muitas questões relevantes surgem quando exploramos o terreno dos métodos mistos na avaliação em suas diversas perspectivas. Os profissionais que seguem diferentes estruturas filosóficas provavelmente responderão de maneira diferente ao uso dos métodos mistos. Por exemplo: os profissionais precisam decidir o momento de usar os métodos mistos, de integrar a análise de dados e de lidar com resultados divergentes.

Além disso, é necessário empreender uma reflexão crítica para decidir quando os métodos mistos são a melhor estratégia, em substituição a métodos unicistas ou múltiplos. Dada a natureza dinâmica do conhecimento, o envolvimento dos interessados e o rápido desenvolvimento dos métodos mistos na história recente da avaliação, esperamos que a comunidade avaliativa continue a contribuir com o avanço da compreensão sobre os métodos mistos, em diversos níveis.

BIBLIOGRAFIA

- ALKIN, M. C. (Ed.). **Evaluation roots: tracing theorists’ views and influences**. Thousand Oaks, CA: Sage, 2004.
- ALKIN, M. C. (Ed.). **Evaluation roots**. 2. ed. Thousand Oaks, CA: Sage, 2013
- BAMBERGER, M. Introduction to mixed methods in impact evaluation. In: **Interaction impact evaluation notes n. 3**. Disponível em: <<http://www.interaction.org/document/guidance>

note-3-introduction-mixed-methods-impact-evaluation>
(último acesso em setembro de 2013).

- BAMBERGER, M.; SEGONE, M. **How to design and manage equity-focused evaluations**. UNICEF: New York, 2012.
- BELL, V. **The Cochrane qualitative methods group**. 2006. Disponível em: <<http://www.lancs.ac.uk/fass/ihr/research/public/cochrane.htm>> (último acesso em setembro de 2013).
- BLEDSON, K. A transformative mixed methods evaluation of an obesity reduction program. In: **Presentation at the annual meeting of the American Evaluation Association**, November 2010. Anaheim, CA.
- BRADY, H. E. Models of causal inference: Going beyond the Neyman-Rubin-Holland Theory. In: **Annual meetings of the political methodology group**. University of Washington, Seattle, Washington. Disponível em: <<http://www.polmeth.wustl.edu/media/Paper/brady02.pdf>> (último acesso em setembro de 2013).
- BRIGGS, J. Cochrane qualitative research methods group. In: **The Cochrane collaboration**. 2006. Disponível em: <<http://www.joannabriggs.edu.au/cqrmg/role.html>> (último acesso em setembro de 2013).
- BRYMAN, A. Integrating quantitative and qualitative research: how is it done? In: **Qualitative research**, 6(1), pp. 97-113, 2006.
- CASEY, K.; GLENNISTER, R.; MIGUEL, E. **The GoBifo project evaluation report: assessing the impacts of community driven development in Sierra Leone**. World Bank, 2011.
- CHILISA, B. **Indigenous research methodologies**. Thousand Oaks, CA: Sage, 2012.
- _____. Education research within postcolonial Africa: a critique of HIV/AIDS research in Botswana. In: **International journal of qualitative studies in education**, 18, pp. 659-684, 2005.
- CHRISTENSEN, C.; MIKKELSEN, M. R.; SICK NIELSEN, T.A.; HARDER, H. Children, mobility, and space: using GPS and mobile phone technologies in ethnographic research. In: **Journal of mixed methods research**, 5, pp. 227-246, 2011. DOI: 10.1177/1558689811406121.
- CRAM, F. Maintaining indigenous voices. In: MERTENS, D. M.;

- GINSBERG, P. E. (Eds.). **The Sage handbook of social research ethics**. Thousand Oaks, CA: Sage, 2009.
- CRESWELL, J. W.; FETTERS, M. D.; PLANO CLARK, V. L.; MORALES, A. Mixed methods intervention trials. In: ANDREW, S.; HALCOMB, L. (Eds.). **Mixed methods research for nursing and the health sciences**. Oxford, UK: Blackwell, 2009.
- DAVIDSON, E. J. "Process values" and "deep values" in evaluation. In: **Journal of multidisciplinary evaluation**, 6(13), pp. 206-208, 2010.
- _____. Book review: evaluation roots: tracing theorists' views and influences. In: **American Journal of Evaluation**, 2, pp. 273-276, 2006.
- DECUIR-GUNBY, J.T.; MARSHALL, P.L.; MCCULLOCH, A.W. Using mixed methods to analyze video data: a mathematics teacher professional development example. In: **Journal of mixed methods research**, 6: pp. 199-216, 2011. DOI: 10.1177/1558689811421174.
- DENZIN, N. K. Triangulation 2.0. In: **Journal of mixed methods research**, 2012.
- _____. Moments, mixed methods and paradigm dialogs. In: **Qualitative inquiry**, 2010. 16 (6), pp. 419-427, 2010. DOI: 10.1177/1077800410364608.
- _____. **The research act: a theoretical introduction to sociological methods**. 3. ed. Englewood Cliffs, NJ: Prentice Hall, 1989.
- DFID. Department for International Development. **The working paper on broadening the range of designs and methods for impact evaluation**. London: Author., 2012.
- GARBARINO, S.; HOLLAND, J. **Quantitative and qualitative methods in impact evaluation and measuring results**. DFID Issue Paper: London, 2009. Disponível em: <<http://www.gsdr.org/docs/open/EIRS4.pdf>> (último acesso em setembro de 2013).
- GERTLER, P.; MARTINEZ, S.; PREMAND, P.; RAWLINGS, L. B.; VERMEERSCH C.M.J. **Impact evaluation in practice**. World Bank Publications, 2011.
- GREENE, J. C. Engaging critical issues in social inquiry by mixing methods. In: **American behavioral scientist**, 56(6), pp. 755-773, 2012.
- _____. Synthesis: a reprise on mixing methods. In: T. S. WEISNER (Ed.). **Discovering successful pathways in children's**

- development: mixed methods in the study of childhood and family life**, pp. 405-420. Chicago, IL: University of Chicago Press, 2005.
- GUBA, E.; LINCOLN, Y. S. **Fourth generation evaluation**. Newbury Park, CA: Sage, 1989.
- HALL, J. Introduction to mixed methods research for nursing and the health sciences. In: ANDREW, S.; E. J. HALCOMB (Eds.). **Mixed methods research for nursing and the health sciences**, pp. 3-12. Oxford, UK: Blackwell, 2009.
- HESSE-BIBER, S. Thinking outside the randomized controlled trials experimental box: strategies for enhancing credibility and social justice. In: MERTENS, D.M.; HESSE-BIBER, S. (Eds.). **New directions in evaluation**, 138, Summer, pp. 49-60, 2013.
- HESSE-BIBER, S. N.; LEAVY, P. Introduction: pushing on the methodological boundaries: the growing need for emergent methods within and across the disciplines. In: HESSE-BIBER, S. N.; LEAVY, P. (Eds.). **Handbook of emergent methods**, pp. 1-15. New York: Guilford, 2008.
- HESSE-BIBER, S. Qualitative approaches to mixed methods practice. In: **Qualitative inquiry**, 16: pp. 455-468, 2010.
- HOWE, K. R. A critique of experimentalism. In: **Qualitative Inquiry**, 10, pp. 42-61, 2004.
- IEG. World Bank and IFC Support for Youth Employment Programs. Washington, D.C.: World Bank, 2012.
- JACKSON, E.T.; TARSILLA, M. Mixed methods in social accounting: evaluating the Micro-Loan Program of Alterna Savings Credit. In: Mook, L. (Ed.). **Accounting for social value**, pp.117-138. Toronto, ON: University of Toronto Press, 2013.
- JOHNSON, R. B.; STEFURAK, T. Dialectical pluralism: a process philosophy for understanding, valuing, and “dynamically combining” our differences. In: MERTENS, D. M.; HESSE-BIBER, S. (Eds.). **New directions in evaluation**, 138, Summer, pp. 37-48, 2013.
- JOHNSON, R. B. **Dialectical pluralism: a metaparadigm to help us hear and “combine” our valued differences**. Paper presented at the Seventh International Congress of Qualitative Inquiry, University of Illinois, Urbana-Champaign, May 21, 2011.
- LUYT, R. A framework for mixing methods in quantitative measu-

- rement development, validation and revision: a case study. In: **Journal of mixed methods research** 6 (4), pp. 294-316, 2011.
- MARK, M.; FEKV, I.; BUTTON, S.B. Integrating qualitative methods in a predominantly quantitative evaluation: a case study and some reflections. In: **New directions for evaluation**, 74, pp. 47-60, 1997.
- MERRIAM, S. **Qualitative research**. San Francisco, CA: Jossey-Bass, 2009.
- MERTENS, D. M. **Transformative research and evaluation**. NY: Guilford, 2009.
- _____. **Research and evaluation in education and psychology: integrating diversity with quantitative, qualitative, and mixed methods**. 3. ed. Thousand Oaks, CA: Sage, 2013.
- MERTENS, D. M.; WILSON, A. T. **Program evaluation theory and practice: a comprehensive guide**. NY: Guilford, 2013.
- MERTENS, D. M.; HESSE-BIBER, S. Mixed methods and credibility of evidence in evaluation. In: MERTENS, D. M.; HESSE-BIBER, S. (Eds.). **New Directions in Evaluation**, 138, Summer, pp. 5-14, 2013.
- MCGEE, R. Qualitative and quantitative poverty appraisal workshop: some reflections and responses. In: KANBUR, R. (Ed.). **Q-Squared: a qualitative and quantitative methods of poverty appraisal**. Permanent Black: Delhi, 2003.
- MOOK, L.; QUARTER, J.; RICHMOND, B. J. **What counts: social accounting for non-profits and cooperatives**. 2. ed. London: Sigel Press, 2007.
- MORSE, J. M. Principles of mixed methods and multi-method research design. In: TASHAKKORI, A.; TEDDLIE, C. (Eds.). **Handbook of mixed methods in social e behavioral research**. pp. 189-208. Thousand Oaks, CA: Sage, 2003.
- NATIONAL INSTITUTE OF HEALTH. **Best practices for mixed methods research in the health sciences**. 2011. Disponível em: <http://obsr.od.nih.gov/mixed_methods_research/pdf/Best_Practices_for_Mixed_Methods_Research.pdf> (último acesso em setembro de 2013).
- NATIONAL SCIENCE FOUNDATION. **User-friendly handbook for**

- mixed method evaluations.** 1997. Disponível em: <<http://www.nsf.gov/pubs/1997/nsf97153/>> (último acesso em setembro de 2013).
- OAKLEY, A. Who's afraid of the randomized controlled trial? Some dilemmas of the scientific method and good research practice. In: OAKLEY, A. **The Ann Oakley reader**, pp. 233-244. Cambridge, UK: Polity, 2005a.
- OAKLEY, A. Paradigm wars: Some thoughts on a personal and public trajectory. In: A. OAKLEY. **The Ann Oakley reader**, pp. 245-250. Cambridge, UK: Polity, 2005b.
- O'CATHAIN, A. Mixed methods research in the health sciences: A quiet revolution [Editorial]. In: **Journal of mixed methods research**, 3, pp. 3-6, 2009.
- OLATZ LOPEZ-FERNANDEZ, O.; MOLINA-AZORIN, J. F. The use of mixed methods research in the field of behavioural sciences. In: **Quant-Qual**, 45: pp. 1459-1472, 2011. DOI 10.1007/s11135-011-9543-9.
- PARLETT, M. & HAMILTON, D. Evaluation as illumination: A new approach to the study of innovatory programs. In: HAMILTON, D.; JENKINS, D.; KING, C.; MACDONALD, B.; PARLETT, M. (Eds.). **Beyond the numbers game: a reader in educational evaluation**, pp. 6-22. London, England: Macmillan, 1977.
- PAWSON, R. Causality for beginners. In: **NCRM Research Methods Festival**, 2008, Leeds University. Disponível em: <<http://eprints.ncrm.ac.uk/245>> (último acesso em setembro de 2013).
- PETTIGREW, M.; ROBERTS, H. **Systematic reviews in the social sciences**. Oxford, England: Blackwell, 2006.
- ROBINSON, A. L.; EMDEN, C. G.; CROFT, T. D.; VOSPER, G. C.; ELDER, J. A.; SIIRLING, C. & VICKERS, J. C. Mixed methods data collection in dementia research: A "progressive engagement" approach. In: **Journal of mixed methods research**, 5(4), pp. 330-344, 2011.
- ROGERS, P.; DAVIDSON, E. J. Australian and New Zealand Evaluation Theorists. In: ALKIN, M. (Ed.). **Evaluation roots**. 2. ed. Thousand Oaks, CA: Sage, 2013.
- SCRIVEN, M. **Evaluation thesaurus**. 4. ed. Thousand Oaks, CA: Sage, 1991.

- SHADISH, W. Evaluation theory is who we are. In: **American Journal of Evaluation**, 19, pp. 1-19, 1998.
- STAME, N. A european evaluation theory tree. In: ALKIN, M. (Ed.). **Evaluation roots**. 2. ed. Thousand Oaks, CA: Sage, 2013.
- STERN, E.; STAME, N.; MAYNE, J.; FORSS, K.; DAVIES, R.; BEFANI, B. DFID Working Paper 38. In: **Broadening the range of designs and methods for impact evaluations**. DFID, London, UK, 2012.
- SWEETMAN, D.; BADJEE, M.; CRESWELL, J.W. Use of the transformative framework in mixed methods studies. In: **Qualitative inquiry**, 16 (6), pp. 441-454. DOI: 10.1177/1077800410364610.
- TARSILLA, M. **Social impact evaluation of the Alterna Savings Community Micro-Loan Program**. Ottawa: Carleton Centre for Community Innovation, Carleton University, 2010.
- TEDDLIE, C.; TASHAKKORI, A. Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In: TASHAKKORI, A.; TEDDLIE, C. (Eds.). **Handbook of mixed methods in social & behavioral research**, pp. 3-50. Thousand Oaks, CA: Sage, 2003.
- TETI, M.; BOWLEG, L.; COLE, R.; LLOYD, L.; RUBINSTEIN, S.; SPENCER, S.; AARON, R.; RICKSECKER, A.; BERHANE, Z.; GOLD, M. A mixed methods evaluation of the effect of the protect and respect intervention on the condom use and disclosure practices of women living with HIV/AIDS. In: **AIDS Behavior**, 14: pp. 567-579, 2010.
- WEHIPEIHANA, N.; DAVIDSON, E. J.; MCKEGG, K.; SHANKER, V. What does it take to do evaluation in communities and cultural contexts other than our own? In: **Journal of multidisciplinary evaluation**, 16(3), pp. 182-192, 2010.
- WHITE, H. The use of mixed methods in randomized controlled trials. In: MERTENS, D. M.; HESSE-BIBER, S. (Eds.). **New directions in evaluation**, 138, Summer, 61-74, 2013.
- _____. **Will we ever learn? Improving lives through impact evaluation**. Washington DC: Center for Global Development, 2006.
- YIN, R. K. **Case study research: design and methods**. 3. ed. Thousand Oaks CA: Sage, 2003.

Vantagens e complexidades das avaliações de impacto

MIGUEL SZÉKELY

Diretor do Instituto de Educação Tecnológica de Monterrey, México.

1. INTRODUÇÃO

Ao longo do século XX, o campo das políticas e programas sociais foi palco de um deslocamento progressivo do foco nos *gastos* como medidas de sucesso para um enfoque concentrado nos *indicadores de insumos* (*inputs*), ou seja, nos bens e serviços considerados determinantes do bem-estar de uma sociedade – como, por exemplo, a quantidade de estradas, escolas, postos de saúde ou habitações disponíveis. Foi apenas em período relativamente mais recente que o interesse migrou dos gastos e indicadores de insumos para os *resultados*, entendidos como evidências baseadas nos indicadores diretos do bem-estar coletivo, tais como bem-estar familiar, capacidade de gerar renda, acumulação de capital humano, condições de saúde, acesso efetivo aos bens e serviços básicos e melhores oportunidades de trabalho, entre outros. A evolução para o foco em resultados implica utilizar outros mecanismos para medir e avaliar o progresso. Enquanto os gastos e insumos podem ser monitorados por meio de mecanismos de auditoria e contabilidade, a mensuração de resultados é geralmente mais complicada, uma vez que demanda novos esforços de coleta de dados, ferramentas

estatísticas e métodos que permitam associar ações específicas de políticas ou programas a impactos também específicos.

Essa mudança de enfoque foi fortemente influenciada pela percepção de diversos agentes vinculados ao financiamento e à ajuda internacional, segundo os quais a quantidade crescente de recursos alocados não impacta necessariamente os níveis de desenvolvimento ou as perspectivas dos países mais pobres – em parte devido ao já mencionado enfoque nos gastos e indicadores de insumos. Birdsall e Savedoff (2010) mostraram que a ajuda externa aumentou cerca de 70% apenas durante a década de 2000 – passando de aproximadamente US\$ 70 bilhões para cerca de US\$ 120 bilhões entre 2000 e 2007 –, ao passo que as preocupações com sua efetividade cresceram paralelamente, em grande parte em função das escassas evidências quanto ao seu efeito nos resultados finais de desenvolvimento¹.

Paradoxalmente, essa mudança relativamente recente de enfoque foi acompanhada, na literatura, por duas preocupações sobre desenvolvimento e políticas sociais. Por um lado, segundo Savedoff, Levine e Birdsall (2006), entre outros autores, porque a quantidade de evidências e conhecimento sobre os efeitos dos diferentes tipos de políticas de intervenções nos resultados do desenvolvimento ainda é pequena. Por outro, existem também percepções documentadas de que as evidências disponíveis são usadas de maneira ainda muito limitada no desenho e execução de tais políticas para lograr seu aperfeiçoamento (vide, por exemplo, Ravallion (2008), Jones et al. (2009) e Weiss (1999)).

Discutiremos neste artigo alguns dos principais problemas que surgem ao se introduzir e implementar avaliações de impacto usando a constituição de grupos aleatórios como dispositivo para medir resultados, com o objetivo de melhorar a prestação de contas e a transparência e também para gerar informações destinadas ao aperfeiçoamento contínuo do desenho de políticas e programas. Como costuma acontecer em todas as ciências sociais, o problema subjacente é que lidar com o comportamento individual

1. Ver Banerjee (2007) e Pitman et al. (2005), para manifestações da mesma preocupação.

como centro das análises, ao avaliar o impacto de determinada ação, implica altos graus de incerteza, e é algo muito mais complexo do que simplesmente medir gastos ou insumos. A mensuração dos resultados requer compreensão do comportamento e das reações do indivíduo, e, embora se tenha avançado substancialmente em termos metodológicos nas últimas décadas, até mesmo a literatura sobre o tema tem sido capaz de oferecer não mais do que aproximações a respeito da avaliação do efeito das ações².

2. AS COMPLEXIDADES DA AVALIAÇÃO DO IMPACTO

Há pelo menos quatro aspectos mais amplos que aumentam a complexidade da mensuração do impacto das políticas sobre o desenvolvimento humano e tornam sua produção e uso extremamente desafiadores:

2.1. HETEROGENEIDADE

O primeiro problema é a heterogeneidade. A ação das políticas geralmente se dirige a grupos de indivíduos que compartilham certas características ou problemas, e raras vezes a ação do governo pode ser concebida para cada circunstância e necessidade individual. No entanto, dependendo das preferências e valores de cada indivíduo e de sua exposição a condições, contextos, situações e antecedentes particulares, a mesma política pode desencadear reações e efeitos totalmente diferentes. Dado que as preferências individuais e os valores inerentes às posições e funções sociais são

2. Não é esse o caso das ciências que analisam materiais ou corpos inertes, nas quais é possível obter baixas margens de erro por meio de avaliações e análises sobre como cada tipo irá reagir ao ser exposto a condições diferentes. As margens de incerteza na política social, em geral, são também muito maiores do que na análise científica em áreas como a saúde, em que previsões podem ser feitas, por exemplo, quanto à reação mecânica do corpo humano diante de um tratamento externo. Nesse caso, haverá sempre margens de erro devido à falta de conhecimento de todas as reações ou mecanismos de resposta possíveis. Porém, uma vez que se estabelece certa regularidade, as variações provavelmente serão decorrentes de elementos externos que não foram até então considerados, e não em função de que um corpo tenha “decidido” subitamente se comportar de modo diferente.

desconhecidos e não passíveis de observação devido a sua heterogeneidade, haverá sempre uma margem de erro e de incerteza em qualquer tentativa de medir as reações a determinada intervenção. Existem abordagens diferentes para se inferir preferências e valores, mas no fim das contas o melhor que se pode esperar é uma boa aproximação baseada em médias, nas quais o valor mediano de uma intervenção talvez não corresponda ao caso de qualquer um dos indivíduos da população³.

2.2. CONTEXTO

O segundo desafio relaciona-se ao contexto. Até mesmo na hipótese de haver informação disponível suficiente a respeito das preferências individuais, elas interagem com condições, contextos, situações e antecedentes diversos, gerando uma resposta específica dos indivíduos às circunstâncias. A informação gerada ao se analisar tais respostas será útil para medir o impacto da ação, porém pode não ser necessariamente relevante para inferir o impacto em outro ambiente. Dado que é praticamente impossível replicar exatamente o mesmo cenário em todos os casos, efeitos idênticos não serão necessariamente observados em uma população diferente. Assim, o que funciona em um caso pode não funcionar em outro. Um exemplo é o efeito da concessão de bolsas de estudos vinculadas à frequência das meninas na escola. Em ambientes onde existir uma preferência cultural pela igualdade de gêneros e participação de mulheres no mercado de trabalho, um apoio em dinheiro poderá ter efeitos significativos na educação, ao passo que exatamente o mesmo apoio, em um ambiente onde se valoriza o casamento precoce, poderia gerar um impacto muito menor.⁴

Contextos diferentes podem levar a uma ampla faixa de efeitos esperados e inesperados, que podem ir além do esforço de

3. Vide Heckman (1997) para uma discussão detalhada sobre os desafios metodológicos que implica a heterogeneidade na avaliação de um programa.

4. Attanasio et al. (2004) apresentam uma abordagem para simular o efeito de transpor para outros cenários o Programa Oportunidades, iniciativa mexicana de transferência de renda, e discutem os desafios e limitações envolvidos nesse tipo de análise.

mensuração e avaliação. As avaliações em geral se preocupam em medir o impacto previsível de uma intervenção e podem até coletar informações adicionais quanto a outras consequências correlacionadas, mas, a menos que haja uma situação de total equilíbrio, onde todos os efeitos cruzados, secundários e adicionais possam ser captados, a informação completa sobre o impacto total não estará disponível. Nesse caso, também podem existir boas aproximações para quantificar consequências previsíveis, mas ainda assim a incerteza quanto a outros impactos poderá deixar abertas enormes avenidas de risco, e o efeito geral da intervenção continuará desconhecido.

Vejamos, por exemplo, o caso do investimento em educação. Ao promover o acúmulo de capital humano, seria de se esperar que as novas gerações tivessem oportunidades de melhorar sua renda, e é possível que assim seja. No entanto, se o contexto estiver moldado por determinadas condições do mercado de trabalho, a incorporação de indivíduos de maior nível educacional leva a um menor retorno associado à escolaridade, e a salários reduzidos, com o impacto final da intervenção podendo ser até negativo na dimensão da renda. Essa é apenas uma ilustração do fato de que nenhuma avaliação terá condições de capturar absolutamente todas as reações desencadeadas por determinada política, e o máximo que podemos esperar é gerar alguma informação quanto à diversidade de dinâmicas que podem ser contempladas.

2.3. MULTIDIMENSIONALIDADE

Outro exemplo dentro da mesma linha tem a ver com a multidimensionalidade. Dado que o bem-estar se expressa por meio de diferentes dimensões, efeitos inesperados poderão ser observados porque afetam este mesmo bem-estar a partir de um ângulo não previsível. Por exemplo: recentemente, programas de transferência condicional de verbas incorporaram o empoderamento das mulheres, dando à chefe de família o controle da transferência de renda pela frequência escolar dos filhos, compromissos de saúde e de nutrição. O argumento é que isso melhora a alocação de recursos no domicílio e melhora as oportunidades das mulheres. Há

de fato evidências de que isso pode ocorrer. Porém, em alguns ambientes, a pessoa que recebe esse novo recurso pode estar sujeita a violência e até a maior discriminação, situações que contrabalançam o efeito da intervenção, resultando em condições negativas ainda piores do que as observadas originalmente. As avaliações que se concentram apenas no efeito direto da transferência de renda poderiam levar a conclusões enganosas – no sentido de inferir que todos os indivíduos daquele lar estão em melhor situação –, devido ao escopo limitado da análise, que não leva em conta o contexto cultural.

2.4. TEMPORALIDADE

Outro aspecto é que a maioria das intervenções para o desenvolvimento tem efeitos intertemporais que não podem ser observados imediatamente. A educação é um bom exemplo. Um melhor capital humano deveria gerar melhores oportunidades, que se traduziriam em mais riqueza material, desenvolvimento profissional, liberdade, cultura etc. Porém, todos esses elementos só poderão ser inteiramente apreendidos se houver um acompanhamento de todos os ciclos de vida de cada indivíduo, assim como dos efeitos sobre o restante da sociedade. Boas aproximações, como fluxos correntes e passados de receitas, situação no mercado de trabalho etc. poderiam trazer alguma informação sobre a relação entre incremento educacional e bem-estar, porém serão sempre mensurações parciais. Outros efeitos positivos, tais como a redução na taxa de criminalidade, impactos sobre a produtividade ou a redução da corrupção, apenas para mencionar algumas áreas, não são geralmente capturados, trazendo como consequência a potencial subestimação dos benefícios da intervenção.

2.5. IDENTIFICAÇÃO

Outra característica é que, mesmo se fosse possível medir efeitos globais, as avaliações raramente têm condições de identificar os mecanismos precisos mediante os quais o resultado é gerado. Todo bem ou serviço têm uma função de produção envolvendo

uma mescla de insumos que podem render efeitos positivos ou negativos e identificar o papel de cada um é virtualmente impossível. Poderia inclusive haver situações em que uma intervenção sublinha o efeito positivo de algum *input* às custas de uma influência negativa sobre outro. Essas dinâmicas, porém, normalmente ficam ocultas no processo de agregação do impacto final. A dificuldade de identificar cada contribuição individual poderá obscurecer o quadro resultante e levar a conclusões equivocadas ou a impressões enganosas quanto à influência dos diferentes elementos.

Quando muitos indivíduos participam da entrega de um serviço ou da operação de um programa, o problema poderá ser ainda maior. Um exemplo são os serviços educacionais. Mesmo passando apenas pela educação primária, estamos expostos a um número grande de professores, materiais, equipamentos, conteúdos etc. que se acumulam até chegar ao nível de capital humano atingido na graduação. A identificação do efeito de cada um dos *inputs* individuais exigiria um acompanhamento do que ocorre praticamente a cada dia com cada um deles, o que é inviável. Isso, acrescido ao caráter intertemporal da acumulação de capital humano, implica, por exemplo, que o efeito de um bom ou mau professor será difícil de identificar no presente – e medir seu efeito total e influência no futuro é impensável.

Da mesma forma, isolar a causalidade entre determinada ação e o efeito sobre um indivíduo ou grupo é também tarefa difícil. Os programas são aplicados em um contexto no qual uma ampla gama de forças e elementos interagem simultaneamente, e a capacidade de extrair uma resposta precisa de uma política específica também pode ser virtualmente impossível. Até mesmo em casos em que se observa uma melhoria das condições de vida depois de implementado determinado programa, não é uma relação direta atribuir o efeito à ação, a menos que o único elemento modificador do *status quo* naquele momento tenha sido exatamente aquela ação específica. Porém, mesmo nesses casos, se houver diferenças nos antecedentes ou nas condições iniciais que provoquem um resultado observado exatamente no mesmo momento em que é esperada a resposta à política – mas que não

tenham relação com o programa –, o efeito poderá ser erroneamente atribuído à intervenção.

Existem diversas formas de lidar com esse problema. Por exemplo: a literatura sobre avaliação de programas sociais tomou emprestadas metodologias consistentes da área médica e de outros campos científicos para desenvolver desenhos experimentais comparando a evolução de um tratamento com grupos controle. Essa abordagem tem proliferado, especialmente para medir o impacto de programas de transferência de renda – o que vem a ser uma evolução considerável com relação a inferir o impacto dos programas na base de comparações amplas do tipo “antes e depois”⁵. É amplamente reconhecido que os desenhos experimentais são capazes de gerar, talvez, as aproximações mais úteis e robustas quanto ao impacto, não estando estas, contudo, isentas de limitações. Além das considerações metodológicas (discutidas, por exemplo, em Deaton (2006)), há desafios técnicos significativos relativos às técnicas econométricas utilizadas (alguns deles resumidos em Ravallion (2008) e Jones et al (2009)). Há, ainda, as considerações éticas e políticas (discutidas mais adiante) que não permitem uma aplicação direta e seu uso generalizado.

Assim, a avaliação da efetividade do desenvolvimento não é fácil e, até mesmo quando se utilizam as melhores ferramentas disponíveis, as evidências geradas são inevitavelmente limitadas. Por um lado, isso, é claro, clama por esforços sustentados para melhorar os dados e os métodos de estimativa. Por outro, exige que seus usuários se mantenham alertas quanto ao potencial e aos possíveis senões das evidências. Um terceiro elemento, que poderia ajudar a explicar o atual *status quo* da baixa produção e do baixo uso dos conhecimentos disponíveis, tem a ver com a natureza dos atores que participam do processo. A seguir, analisa-se esse aspecto.

5. Skoufias (2005) documentou o que talvez tenha sido o primeiro programa de transferência de renda avaliado usando técnicas que vão além da mensuração de gastos, *inputs* ou simplesmente pela comparação da situação antes e depois da implementação. O programa Progreso (sigla em espanhol do programa de educação, alimentação e saúde), iniciado em 1998 no México, foi acompanhado desde o início por uma avaliação com desenho experimental.

3. A IMPORTÂNCIA DE COMPREENDER OS ATORES ENVOLVIDOS

Analisaremos brevemente o papel dos atores que são chave no processo de avaliação do impacto.

3.1. OS INVESTIDORES PRIVADOS E A REALIZAÇÃO DE AVALIAÇÕES DE IMPACTO

Um dos atores que mais influem no processo de produção e uso das avaliações são os doadores ou investidores. Eles são externos à operação do programa e à agência encarregada de sua execução, mas proporcionam o financiamento para a implementação das políticas e ações. São desde instituições e fundações multilaterais até entidades privadas em âmbito local, cujo interesse é apoiar esforços destinados a resolver determinado problema. Existem diversos motivos para que esses atores não sejam, conscientemente, partidários da mensuração do impacto da aplicação de seus recursos sobre os resultados do desenvolvimento e que até, em algumas circunstâncias, minem, racionalmente, os esforços de avaliação.

O primeiro motivo é que avaliar exige recursos. Para realizar um processo adequado, normalmente é preciso produzir dados sob medida, investir no desenvolvimento de um desenho adequado e financiar os serviços profissionais de análise de dados e elaboração dos relatórios de achados e conclusões. Quando os recursos são limitados (e esse é frequentemente o caso), surge o dilema moral de destinar mais recursos para a avaliação do impacto ou para beneficiar mais pessoas. O argumento óbvio a favor do primeiro é que a avaliação ajudará a verificar se a ação teve impacto, positivo ou negativo, e que seus resultados evidenciarão como melhorar a efetividade e a eficácia dos recursos gastos, com maiores benefícios para uma população mais ampla no futuro. O argumento contra a alocação de recursos para essa atividade surge quando as necessidades são tão graves que qualquer recurso que não seja destinado a um beneficiário poderia significar a diferença entre a vida e a morte de um ser humano.

O segundo motivo é que, sendo o conhecimento um bem público, o incentivo para alocar recursos para a avaliação, com benefícios que vão muito além do interesse imediato de conhecer

melhor o impacto de uma ação, é muito menor. Esse argumento é debatido em detalhe por Avery et al (1999) e Savedoff et al (2006). O aspecto de bem público das avaliações é exacerbado quando existem múltiplos doadores participando da arrecadação de fundos para grandes iniciativas, o que faz que cada um perceba a sua contribuição como meramente marginal (“uma gota no oceano”). Sob essa ótica, o incentivo e o poder de cada doador para exigir uma prestação de contas são fracos.

3.2. OS FORMULADORES DAS POLÍTICAS

O segundo ator estratégico é o responsável pela formulação das políticas e pelo processo de decisão de alto nível, que pode ser simultaneamente usuário e produtor (indireto) dos resultados da avaliação. É usuário em potencial na medida em que as decisões gerais que orientam o financiamento e a implementação das políticas estão sob sua responsabilidade, e a informação gerada pelas avaliações poderá ser estratégica para identificar as áreas de melhoria e para fornecer *feedback* para tomar decisões de melhor qualidade. Pode ser considerado produtor visto que é sua responsabilidade autorizar a realização de uma avaliação de programa ou atividade pública sob sua jurisdição. Nesse caso, podemos imaginar os motivos a favor ou contra a alocação de recursos para viabilizar uma avaliação.

Uma explicação para o uso limitado das avaliações é que as decisões sobre políticas são tomadas dentro de um quadro de restrições e condições que poderiam desviar o interesse pela atividade. Entre os fatores relevantes destacam-se o tempo, os grupos de interesse, os aspectos normativos, as decisões institucionais, as políticas e o financiamento. A questão está frequentemente em conflito com a dinâmica de produzir evidências confiáveis. Na ação pública, os prazos para entregar resultados são rígidos, enquanto uma avaliação sólida de impacto requer tempo para ser desenhada, captar dados de linha de base, gerar informações *ex-post* e, por fim, produzir o relatório e a análise. Tudo isso poderia significar anos além de um ciclo político, o que desencorajaria até os mais fervorosos adeptos das avaliações, uma vez que estes não teriam condições de colher os frutos da iniciativa.

Os grupos de interesse também podem representar uma significativa restrição para os formuladores de políticas e impedi-los de promover avaliações de impacto, especialmente no caso de desenhos experimentais. A definição dos grupos de tratamento e de controle poderia ser perfeitamente justificável do ponto de vista metodológico; porém explicar aos excluídos dos benefícios que não foram “selecionados ao acaso”, enquanto outros o foram, geraria oposição suficiente para inviabilizar a avaliação. Até mesmo nos casos em que um experimento possa ser lançado, as evidências de resultados positivos poderiam gerar pressões do grupo de controle para ser incorporado, o que é perfeitamente compreensível, ameaçando “contaminar” o experimento. O interessante é que as pressões normalmente surgem quando é mais importante manter o grupo de controle como tal, uma vez que o efeito pleno pretendido só poderá ser medido quando a intervenção tiver tido tempo de apresentar seus impactos de médio ou longo prazos. A capacidade dos formuladores para lidar com as pressões políticas ou dos grupos de interesse é, em geral, o fator determinante da viabilidade do experimento.

Outro aspecto importante são os fatores normativos. Se, por exemplo, as regras e as regulamentações impedem o uso dos recursos do programa para avaliar o seu impacto, mesmo quando os formuladores da política estão a favor da avaliação, haverá um subinvestimento nessa atividade. Da mesma forma, as disposições institucionais poderiam ser um obstáculo. Quando os governos são alinhados em torno do conceito de medir gastos e insumos, pode não existir um mandato para avaliar resultados ou sequer agências capacitadas a fazê-lo, tornando impossível ou mesmo ilegal a realização da avaliação.

E há ainda a questão política. Gerar informações sobre a eficiência da ação de determinada iniciativa pode representar um alto risco em certos ambientes. Enquanto os efeitos positivos podem ser capitalizados politicamente, os resultados desfavoráveis tendem a ser muito mais difíceis de encarar, exigindo investimentos de capital político para lidar com eles. Como argumenta Pritchett (2002), o risco de obter resultados negativos (ou não tão positivos) com uma avaliação pode ser um forte impedimento

para realizá-la. Os resultados dariam munição para a oposição, com resultados politicamente adversos e até letais, ou desencorajariam doadores e investidores externos a contribuir com mais recursos. Normalmente, o risco é mais alto em sociedades onde os mecanismos de transparência e prestação de contas são institucionalizados e em ambientes de restrições orçamentárias agudas (onde, paradoxalmente, ter informação sobre quais políticas são as mais necessárias é muito importante), nos quais diversos grupos de interesse concorrem pelos recursos. Nessas circunstâncias, fornecer evidências sólidas quanto ao impacto de um programa pode ser o equivalente a assinar seu atestado de óbito.

3.3. OS AVALIADORES

O avaliador tem um papel importante no processo. É comum que este seja uma pessoa externa à operação do programa e que tenha independência suficiente para garantir credibilidade. As autoavaliações também são uma alternativa, mas como estão atravessadas pela subjetividade, sua credibilidade fica prejudicada. Como foi reconhecido, pelo menos desde Alkin e Daillak (1979), a abordagem escolhida pelos avaliadores é variável crítica para promover ou desencorajar o uso e a produção de evidências. Do lado positivo, gerar informação sólida e confiável quanto aos efeitos da ação de uma política pode ser de grande valor para os doadores e investidores externos que precisam se certificar sobre o efeito de sua intervenção no desenvolvimento. Para os formuladores de intervenções, a informação oportuna, confiável e relevante, que ajuda a capitalizar ações positivas de governo e que oferece elementos para melhorar o desempenho na esfera política, é algo apreciado. O risco político dos resultados não tão bons pode às vezes ser até mitigado se, desde o início, os formuladores e gestores do programa se envolverem na definição das perguntas a serem respondidas e da estratégia para a comunicação adequada dos resultados.

Por outro lado, a escolha de abordagens que restringem o uso das avaliações poderia prejudicar o interesse dos formuladores das políticas e/ou dos doadores e investidores externos. Por exem-

plo: um estudo que requer esperar uma geração inteira para obter informações relevantes poderá ser inteiramente inviável para ambos. Da mesma forma, as abordagens metodológicas que geram resultados válidos apenas para um determinado ambiente podem ser de escassa utilidade para ampliar a escala de uma determinada ação ou para fundamentar decisões em outros contextos. Concentrar-se em questões acadêmicas interessantes, porém operacionalmente irrelevantes, desencorajará, da mesma maneira, a demanda por avaliações. As conclusões sensatas do ponto de vista metodológico, mas inviáveis no âmbito político, em geral encontram contrapartes pouco receptivas, enquanto as avaliações ideais no campo técnico, mas que exigem orçamentos inalcançáveis ou difíceis de administrar, muito provavelmente não serão realizadas – ou, se o forem, é quase certo que sejam pouco usadas.

Tais situações podem ocorrer porque os objetivos e as restrições dos doadores e investidores externos e dos formuladores das políticas não necessariamente coincidem com os dos profissionais com capacidade para fazer avaliações sólidas. Esses, sobretudo dos círculos acadêmicos, podem priorizar a pureza teórica, o prestígio profissional, o reconhecimento, a geração de conhecimento, o sucesso em termos de publicações em meios ilustres etc., aspectos incompatíveis com avaliações oportunas, confiáveis, relevantes, pertinentes e comunicáveis do ponto de vista dos usuários. Uma das razões para explicar a percepção de que a informação disponível é subutilizada poderia muito bem ser a sua incompatibilidade com as necessidades dos usuários, dentro das linhas descritas anteriormente.

3.4. GESTORES E TÉCNICOS DE PROGRAMAS

Por diversos motivos, talvez os atores mais afetados, em termos práticos, pelo processo de geração de evidências das avaliações e da implementação de mudanças nos programas e serviços segundo seus resultados sejam os gestores das iniciativas e os técnicos que trabalham diretamente no campo. Da perspectiva da geração de evidências, uma das razões é que a produção de dados e de informações necessárias para desenhar e implementar as avaliações

em geral demanda trabalho e esforços adicionais, que não vêm necessariamente acompanhados de incentivos ou de benefícios identificáveis a curto prazo para eles. Além disso, falhar na entrega dos benefícios para as populações específicas, de acordo com os cronogramas rígidos, poderia colocar em risco a avaliação ou contaminar um experimento, com consequências significativas para a qualidade do trabalho. Há ainda outro motivo, no caso de desenhos experimentais: que os técnicos estão diretamente expostos às pressões dos indivíduos do grupo de controle, ou de outros grupos, que exigem ser incorporados ao programa.

Mas talvez o principal desafio se relacione ao uso dos resultados das avaliações. Passar de uma abordagem baseada em gastos ou indicadores de insumos como medida do sucesso da política para enfocar resultados de desenvolvimento como *benchmark* é, em si, uma mudança cultural significativa. Quando as novas avaliações são apresentadas e são propostos conjuntos de recomendações que vão modificar normas, processos e as práticas cotidianas, os que “têm o ônus” de colocar em prática essas mudanças são precisamente os técnicos. A introdução e a implementação gradual de novos padrões e métodos nas operações e nas atividades do dia a dia pode ser a mais complexa e laboriosa tarefa no uso efetivo das avaliações, dado que exige uma mudança de conduta e costumes individuais. É comum que mesmo o que parece uma mudança mínima nas normas e processos leve vários anos para ser plenamente implementado – e a oposição implícita ou explícita nesse nível poderia inibir o uso de avaliações.

Também é frequente que gestores e técnicos se tornem, em alguns casos, o “eleitorado” do programa ao qual dedicaram anos de esforços e experiência. Quando esses indivíduos são também os clientes de seu próprio programa, podem ser os primeiros a obstruir ou a se opor abertamente às mudanças, tornando o uso da avaliação impossível na prática. Mesmo se, no nível decisório mais alto, houver forte compromisso quanto ao uso dos resultados da avaliação para aperfeiçoar o desenho e a implementação das políticas, a resistência na base pode inviabilizar a avaliação. Seja por razões práticas – como exigir esforços adicionais –, seja porque estão intimamente identificados com o programa e com a

maneira como foi desenhado e gerido até então, qualquer desafio ao *status quo* é encarado como algo pessoal.⁶

O treinamento intensivo e a divulgação de informação sobre a natureza e os propósitos das avaliações contribuiriam para reduzir a oposição a sua produção e uso. A experiência tem demonstrado que envolver esses atores no processo de desenho da avaliação (por exemplo: participando na definição das questões a serem respondidas, incluindo-os no processo de elaboração das hipóteses, e até identificando as áreas de melhoria potencial), poderia ajudar a redirecionar seus esforços para um uso mais efetivo dos resultados obtidos com as avaliações.

4. COMENTÁRIO FINAL

Dada a complexidade técnica e os diferentes incentivos aos quais estão sujeitos os atores-chave envolvidos no processo, não é surpresa que as avaliações experimentais nem sempre sejam realizadas, nem mesmo em ambientes em que poderiam representar uma ferramenta estratégica para delinear o desenho de um programa.

Existem evidentemente outras abordagens metodológicas para fazer avaliação, mas seu uso, como ocorre com os métodos experimentais, dependerá do contexto específico, do tempo oportuno e da situação que prevalece no momento em que for necessário gerar evidências sobre os efeitos de um determinado programa ou intervenção.

BIBLIOGRAFIA

ALKIN, M. C.; DAILLAK, R. H. A study of evaluation utilization. In: **Educational Evaluation and Policy Analysis**, v. 1, n. 4, jul.-ago., pp. 41-49, 1979.

ALKIN, M. C.; DAILLAK, R.; WHITE, P. Using evaluations: does eva-

6. Ferman (1969) apresenta uma caracterização interessante do relacionamento geral entre avaliadores e gestores. Embora o trabalho tenha sido publicado há 40 anos, ilustra claramente os conflitos e tensões que podem surgir entre os diferentes atores do processo de avaliação, o que pode ser determinante para sua viabilidade.

- uation make a difference?. In: **Library of Social Research** #76. Beverly Hills: Sage Publications, 1979.
- ANDERSON, O.W. Influence of social and economic research on public policy in the health field: a review. In: **Milbank Memorial Fund Quarterly**, v. 44, n. 3, pp. 11-51, 1966.
- ATTANASIO, O.; MEGHIR, C.; SZÉKELY, M. Using randomized experiments and structural models for “scaling up”: evidence from the Progres evaluation. In: BOURGUIGNON, F.; PLESKOVIC, B. (Eds). **Annual World Bank Conference in Development Economics**. Oxford University Press, 2004.
- AVERY, C.; RESNICK, P.; ZECKHAUSER, R. The market for evaluations. In: **The American Economic Review**, v. 89, n. 3, pp. 564-584. American Economic Association, 1999.
- BANERJEE, A. **Making aid work**. MIT Press, 2007.
- BARBER, M.; MOURSHED, M. Cómo hicieron los sistemas educativos con mejor desempeño del mundo para alcanzar sus objetivos. In: **Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL)**, n. 41, Santiago do Chile, jul. 2008.
- BARKDOLL, G. L. Type III evaluations: consultation and consensus. In: **Public Administration Review**, v. 40, n. 2, mar.-abr., pp. 174-179, 1980.
- BIRDSALL, N.; SAVEDOFF, W. **Cash on delivery**: a new approach to foreign aid. Center for Global Development, 2010.
- BRASKAMP, L. A.; BROWN, R. D. **Utilization of evaluative information**. San Francisco: Jossey-Bass, 1980.
- BROWN R. D.; NEWMAN D. L.; RIVERS L. S. A decision making context model for enhancing evaluation utilization. In: **Educational Evaluation and Policy Analysis**, v. 6, n. 4, pp. 393-400, inverno de 1984.
- CARO, F. G. Issues in the evaluation of social programs. In: **Review of Educational Research**, v. 41, n. 2, Science and Mathematics Education, pp. 87-114, abr. 1971.
- CUMMINGS, R. Rethinking evaluation use. In: **Australasian Evaluation Society**. Nov. 2002.
- DEATON, A. Evidence-based aid must not become the latest in a long string of development fads. In: **Boston Review**. Jul. 2006.
- DRUMMOND, M. Making economic evaluations more accessible

- to health care decision-makers. In: **The European Journal of Health Economics**, v. 4, n. 4, pp. 246-247, 2003.
- FERMAN, L. A. Some perspectives on evaluating social welfare programs. In: **Annals of the American Academy of Political and Social Science**, 1969.
- GAARDER, M. Introduction. In: **Journal of Development Effectiveness**, v. 3, n. 3, 2010.
- GAARDER, M.; GLASSMAN, A.; TODD, J. Conditional cash transfers and health: unpacking the causal chain. In: **Journal of Development Effectiveness**, v. 3, n. 3, 2010.
- GONZALEZ DE COSSIO, T.; RIVERA, J.; LÓPEZ ACEVEDO, G. Nutrición y pobreza. In: **World Bank**, Washington DC, 2008.
- GREENBERG, B. G. Evaluation of social programs. In: **Review of the International Statistical Institute**, 1968.
- HADDAD, W.; DRAXLER, A. **Technologies for Education: potentials, parameters, and prospects**. UNESCO, Paris, 2002.
- HANUSHEK, E.; KAIN, J.; O'BRIEN, D. M.; RIVKIN, S. The market for teacher quality. In: **Working Paper 11154**, National Bureau of Economic Research, Cambridge, Massachusetts, 2005.
- HANUSHEK, E.; WOESSMANN, L. The role of school improvement in economic development. In: **CESifo Working Paper**, n. 1911, fev. 2007.
- HENRY, G.; MARK, M. Beyond use: understanding evaluation's influence on attitudes and actions. In: **American Journal of Evaluation**, v. 24, n.3, pp. 293-314, 2003.
- HECKMAN, J.; SMITH, J.; CLEMENS, N. Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. In: **Review of Economic Studies**, 64, n. 4, pp. 487-535, 1997.
- HOWARD R.; SALASIN, D.; SALASIN, S. The utilization of evaluation. In: **Handbook of Evaluation Research**, v. 1, E. L. Streuning and M. Gutentag, Beverly Hills, CA: Sage, 1975.
- JOHN, E. S. L.; THOMAS, J. C. **Designing useful evaluations, evaluation and program planning**, 5(4):327-36, 1982.
- JONES, N.; JONES, H.; STEER, L.; DATA, A. **Improving impact evaluation production and use**. Overseas Development Institute, 2009.

- JORDAN, H.; MENDRO, R.; WEERASINGHE, D. Teacher effects on longitudinal student achievement. In: **Mimeo**, Indianapolis, IN, 1997.
- JUNG, I. ICT - pedagogy integration in teacher training: application cases worldwide. In: **Educational Technology & Society**, 8 (2), pp. 94-101, 2005.
- KORETZ, D. Developing useful evaluations: a case history and some practical guidelines. In: **New Directions for Program Evaluation: Making Evaluation Research Useful to Congress**, L. Saxe and D. Koretz, San Francisco: Jossey-Bass, pp. 25-50, 1982.
- KIRKHART, K. E. Reconceptualizing evaluation use: an integrated theory of influence. In: CARACELLI, V.J.; PRESKILL, H. (Eds.). **The expanding scope of evaluation use**. New Directions for Evaluation, n. 88, 2000.
- LEONARD, R. "Barriers to the utilization of evaluation research". In: *Annual Meeting of the Society for the Study of Social Problems*, Chicago, IL, set. 1977.
- LEVINTON L. A.; HUGHES, E. F. Research on utilization of evaluations: a review and synthesis. In: **Evaluation Review**, pp. 525-548, 1981.
- MARK, M.; HENRY, G. The mechanisms and outcomes of evaluation influence. In: **Evaluation**, n. 10, 2004.
- MARRA, M. How much does evaluation matter: some examples of the utilization of the evaluation of the World Bank's anti corruption activities. In: **Evaluation**, v. 6, n. 1, pp. 22-36, 2000.
- OECD. **Teachers matter; attracting, developing and retaining effective teachers**. Organization for Economic Co-Operation and Development, Paris, 2005.
- OECD. **Are the new millennium learners making the grade?** Organization for Economic Co-Operation and Development, Centre for Educational Research and Innovation, Paris, 2010.
- PATTON, M., **In search of impact: an analysis of the utilization of federal health evaluation research**. Center for Social Research, Minnesota, 1975.
- PATTON, M. Q. **Utilization-focused evaluation**. Beverly Hills: Sage Publications, 1978.

- PELGRUM, H. Promoting equity through ICT: what can international assessments contribute to help fight low achievement?. In: KARPATI, A. (Ed.). **Promoting equity through ICT in education: projects, problems, prospects**. Budapest, Hungarian Ministry of Education, OECD, 2004.
- PITTMAN, G.; FEINSTEIN, O.; INGRAM, G. **Evaluation development effectiveness**. New Brunwsick, Transaction Publishers, 2005.
- PRITCHETT, L. It pays to be ignorant. In: **Journal of Economic Policy Reform**, v. 5, n. 4, pp. 251-269, 2002.
- RAVALLION, M. **Evaluation in the practice of development**. World Bank, 2008.
- RAWLINGS, L. Evaluating the impact of conditional cash transfer programs. In: **World Bank Research Observer**, v. 20 (1), pp. 29-55, 2005.
- RIVKIN, S. G.; HANUSHEK, E.; KAIN, J. Teachers, schools and academic achievement. **Econometrica**, v. 73, n. 2, pp. 417-458, mar. 2005.
- SANDERS, W.; RIVERS, C. **Cumulative and residual effects of teachers on future student academic achievement**. University of Tenesee, 1996.
- SAVEDOFF, W.; LEVINE, R.; BIRDSALL, N. **When will we ever learn? Improving lives through impact evaluation**. The Evaluation Gap Working Group, Center for Global Development, Washington DC, 2006.
- SADOFSKY, S. Utilization of evaluation results: feedback into the action program. In: SHMELZER, J. (Ed.). **Learning in action**. Washington: U.S. Government Printing Office, 1966.
- SCHULBERG, H.; BAKER, F. Program evaluation models and the implementation of re- search findings. In: **American Journal of Public Health**, 1968.
- SKOUFIAS, E. Progesa and its impact on the welfare of rural households in Mexico. In: **International Food Policy Research**, Washington DC, 2005.
- SWEET, R.; MEATES, A. ICT and low achievers: what does PISA tell us?. In: KARPATI, A. (Ed.). **Promoting equity through ICT in education: projects, problems, prospects**. Budapest, Hungarian Ministry of Education, OECD, 2004.

- SZÉKELY, M. **Números que mueven al mundo: la medición de la pobreza en México**. Editorial Porrúa, México DF, 2006.
- SZÉKELY, M.; RASCÓN, E. México 2000-2002: reducción de la pobreza con estabilidad y expansión de programas sociales. In: **Economía mexicana**, v. XIV, n. 2, 2005.
- WALTER, J. Can evaluations influence programs? The case of compensatory education. In: **Journal of Policy Analysis and Management**, v. 2, n. 2, pp. 174-184, invierno de 1983.
- WEISS, C. H. The many meanings of research utilization. In: **Public Administration Review**, set-out. 1979.
- WEISS, C. H. Measuring the use of evaluations. In: HOUSE, E. (Ed.). **Evaluation studies review annual**, v. 7, pp. 129-146, Beverly Hills, CA: Sage, 1982.
- WEISS, C. H. The interface between evaluation and public policy. In: **Evaluation**, v. 5, n. 4, pp. 468-486, 1999.
- WEISS, C. H. Theory-based evaluation: theories of change for poverty reduction programs. In: FEINSTEIN, O.; PICCIOTO, R. (Eds.). **Evaluation and poverty reduction**, Transaction Publications, 2001.

Avaliações válidas, relevantes e úteis

THOMAZ KAUARK CHIANCA

Doutor em avaliação pela Universidade de Western Michigan, consultor internacional de avaliação e odontólogo da Universidade Federal do Rio de Janeiro.

1. INTRODUÇÃO

O interesse pelo campo da avaliação no Brasil tem aumentado significativamente nos últimos anos. É notável a crescente demanda por avaliações profissionais por parte dos diferentes setores da sociedade – governos, organizações do terceiro setor e empresas. O aumento do número de eventos ligados ao tema e a criação de organizações que reúnem profissionais que atuam na área, como a Rede Brasileira de Monitoramento e Avaliação e a Associação Brasileira de Avaliação Educacional, são exemplos claros desse cenário.

Apesar dos avanços, a avaliação enquanto campo profissional no Brasil ainda está em sua infância. A produção de conhecimento sobre a teoria e a prática da avaliação no contexto brasileiro é ainda limitada, tanto em termos de quantidade quanto de qualidade. No momento, são apenas duas as instituições de ensino em todo o País que oferecem formação específica para profissionais interessados em se tornar avaliadores. Fica clara, portanto, a necessidade de realizar avanços para que a avaliação alcance maior maturidade, que possa ser refletida no aperfeiçoamento das práticas adotadas pelos avaliadores.

Uma preocupação constante por parte das organizações contratantes é com a qualidade das avaliações realizadas. São vários os relatos de frustração com os processos avaliativos que não as

ajudam. Tais avaliações normalmente geram um relatório extenso, com detalhes sobre métodos e instrumentos de pesquisa utilizados, mas portadores de um defeito fatal: não respondem diretamente às perguntas realmente importantes. E, conseqüentemente, não ajudam a melhorar ou a tomar decisões estratégicas sobre o futuro daquilo que estava sendo avaliado, o “avaliando”. O presente artigo propõe-se a contribuir para qualificar o trabalho realizado por avaliadores, principalmente os que atuam no terceiro setor, de modo que as avaliações que realizem tenham mais validade e sejam mais relevantes e úteis.

São três as principais mensagens contidas neste artigo:

1. A avaliação vai muito além da aplicação de métodos de pesquisa para estudar variáveis relacionadas ao “avaliando”, normalmente as que são de medição direta, rápida e confiável. O que muitos profissionais não se dão conta, ou não se perguntam mais seriamente, é se estas variáveis são os aspectos realmente importantes a serem considerados para determinar se o “avaliando” tem qualidade, importância ou valor. É muito fácil ficar perdido em métodos e medições. As avaliações, porém, superam aferição de variáveis. Elas exigem a apresentação de respostas bem fundamentadas para perguntas importantes e realmente avaliativas.
2. Há metodologias que são específicas do campo da avaliação, e que diferenciam efetivamente um processo avaliativo de uma pesquisa. Se não são utilizadas, torna-se impossível identificar e selecionar os parâmetros-chave para determinar quão bom um “avaliando” é, e tampouco chegar a conclusões avaliativas válidas sobre seu desempenho nas dimensões e/ou componentes considerados estratégicos.

1. Para evitar repetições, o termo “avaliando” (entre aspas) será utilizado no decorrer do texto para representar os diferentes tipos de “coisas” que podem ser avaliadas: programas, projetos, políticas, produtos, serviços, organizações, pessoas, performance, propostas, avaliações etc.

3. A essência da avaliação está na resposta válida e direta a perguntas realmente relevantes, que ajudem a tomar decisões sobre a melhoria ou o futuro de um “avaliando”. Para que uma avaliação consiga cumprir essa sua função essencial com qualidade é preciso desprender-se da ideia de reproduzir pura e simplesmente a lógica das pesquisas acadêmicas tradicionais. Há seis elementos-chave que precisam ser considerados com cuidado para realizar avaliações com qualidade. Tais elementos serão apresentados neste artigo.

2. PARA ALÉM DA PESQUISA

Há profissionais que consideram a avaliação nada mais do que a aplicação de métodos de pesquisa tomados emprestados, principalmente das ciências sociais e econômicas, para apoiar decisões ou responder a indagações sobre um determinado “avaliando”. Quantos já não se depararam com a situação representada na tira abaixo? >



Fonte: o autor, utilizando ferramentas do site <www.makebeliefscomix.com>.

Normalmente, a encomenda feita aos avaliadores é para que determinem em que medida os objetivos do programa foram atingidos a partir de alguns indicadores – de preferência quantitativos e de aferição relativamente fácil e consistente. Como ato contínuo à primeira conversa entre o avaliador e o cliente, dá-se um salto direto para escolher quais os métodos de pesquisa que melhor se prestam para identificar o cumprimento ou não dos objetivos traçados, com base nos indicadores selecionados. Indagações bási-

cas sobre, por exemplo, em que medida os objetivos estabelecidos eram realmente válidos ou atendiam às necessidades específicas de pessoas ou organizações; ou quais os atributos essenciais para determinar qualidade, importância ou valor do “avaliando” passam, em geral, ao largo das discussões.

No exemplo (exagerado) representado na ilustração, há questões básicas que um avaliador precisaria levantar para fazer um bom trabalho. Entre vários outros questionamentos que precisam ser feitos para que uma avaliação efetivamente gere conclusões válidas e relevantes, podemos citar: “Será que o número de árvores a serem plantadas é suficiente para resolver o problema do desmatamento na região mencionada?”; “Será que o tipo de árvore plantada era o mais adequado?”; “As árvores foram bem cuidadas e vingaram ao longo do tempo?”; “Há um trabalho de manutenção e proteção dessas árvores?”.

Michael Scriven, um dos mais importantes nomes dessa área de conhecimento em todo mundo, ajuda-nos a identificar a diferença essencial entre a avaliação e a pesquisa. Ele defende que um estudo só pode ser considerado uma avaliação se chegar a conclusões sobre quão bom, importante ou valioso é o “avaliando” (Scriven, 1991). Independentemente do método selecionado para coletar e analisar dados, é preciso que esse gere conclusões sobre a situação ou desempenho daquilo que está sendo estudado. Esse ponto é essencial, pois muitas das críticas que os avaliadores recebem estão baseadas no fato de que grande parte do trabalho que realizam não responde diretamente às principais preocupações daqueles que precisam tomar decisões.

Para tornarem-se bons avaliadores, é necessário que os profissionais desenvolvam uma percepção clara da lógica e dos métodos que formam a essência do processo avaliativo. E o que o torna diferente da pesquisa não avaliativa é a combinação bem fundamentada de valores (por exemplo: necessidades ou aspirações da população-alvo de um programa, critérios éticos, custo-efetividade etc.) que servirá de base para julgar a qualidade, a importância e a *performance* do que está sendo estudado. Infelizmente, os profissionais que atuam nesse campo têm dado menos atenção à essência da avaliação do que à aplicação e adaptação

de métodos de pesquisa quantitativos e qualitativos, deixando de levar em conta que as metodologias específicas desta área de conhecimento são a principal garantia da validade, relevância e dos usos de uma avaliação.

3. MÉTODOS ESPECÍFICOS DA AVALIAÇÃO

Os métodos particulares do campo da avaliação fazem que estudos possam ser considerados efetivamente avaliativos. A seguir serão descrito quatro métodos.

3.1. IDENTIFICAÇÃO E SELEÇÃO DOS “VALORES” DE UMA AVALIAÇÃO

“Valores” são os princípios ou as propriedades inerentes dos “avaliandos” que os tornam bons e relevantes. Em qualquer avaliação, os “valores” formam a base para a identificação e a seleção dos critérios e indicadores que irão orientar a coleta de dados e o estabelecimento de conclusões avaliativas. O atendimento a necessidades, direitos e/ou prioridades das pessoas, grupos, organizações etc., que serão servidos de alguma maneira pelo “avaliando” é a principal fonte de “valores” que o avaliador deve considerar.

Na avaliação de programas, tipo de estudo realizado com maior frequência, o profissional precisa identificar em que medida se está realmente *dando conta de necessidades importantes de pessoas e grupos*. Por exemplo: um programa de ações complementares à escola (realizado no contraturno escolar) pode conseguir atender bem à necessidade de manter as crianças ocupadas e, potencialmente, menos vulneráveis aos riscos das ruas. Porém, pode não contribuir de maneira significativa para a solução de outras necessidades igualmente ou mais importantes – como desenvolver as habilidades cognitivas, sociais e emocionais necessárias para o sucesso na escola e na vida de maneira geral.

Outras fontes determinantes de valores em uma avaliação de programas incluem:

- Alcance de resultados: o programa deverá produzir resultados significativos para o benefício de famílias e comunidades participantes, sem gerar efeitos colaterais negativos significativos.
- Solidez científica: a lógica do programa e suas prioridades, estratégias e atividades devem ser consistentes com o conhecimento mais atual sobre as questões que estão sendo abordadas.
- Padrões éticos e profissionais: o programa deve estar em conformidade com princípios profissionais e éticos aceitos pela sociedade.
- Eficiência: o programa deve permitir o mínimo de desperdício e usar os recursos com eficiência máxima.
- Custo-efetividade: o programa deve produzir os resultados esperados, dentro de padrões éticos e de qualidade, pelo menor custo em comparação com outras opções.

Para cada “avaliando”, os profissionais terão de trabalhar junto com os principais interessados para determinar quais os “valores” específicos que, naquele contexto, deverão ser levados em conta para orientar o processo avaliativo.

3.2. DETERMINAÇÃO DO DESEMPENHO DO “AVALIANDO” EM RELAÇÃO A DIMENSÕES, INDICADORES E/OU COMPONENTES

Após a coleta das diversas evidências quantitativas e qualitativas a partir dos “valores” selecionados, o avaliador precisa determinar como foi o desempenho de um “avaliando” em relação a diferentes dimensões de qualidade e/ou importância. O grande desafio é como combinar diferentes evidências para concluir se o desempenho foi excelente, bom, razoável ou ruim. Uma ferramenta muito efetiva são as “rubricas avaliativas”.

A área de educação foi a primeira a utilizar rubricas como instrumento para orientar os professores a avaliar a qualidade de redações e de outros trabalhos escritos pelos alunos. Mais recentemente, têm sido adotadas por avaliadores na elaboração de parâmetros para determinar o grau de sucesso de um “avaliando” em diferentes dimensões de resultado, processo e/ou contexto. Concretamente, uma rubrica avaliativa apresenta uma descrição clara

dos diferentes níveis de desempenho (ou situação) de um “avaliando”. Ou seja, traz os detalhes que possibilitam aos avaliadores afirmar se o desempenho ou a situação do “avaliando”, em relação às diferentes dimensões analisadas, podem ser considerados:

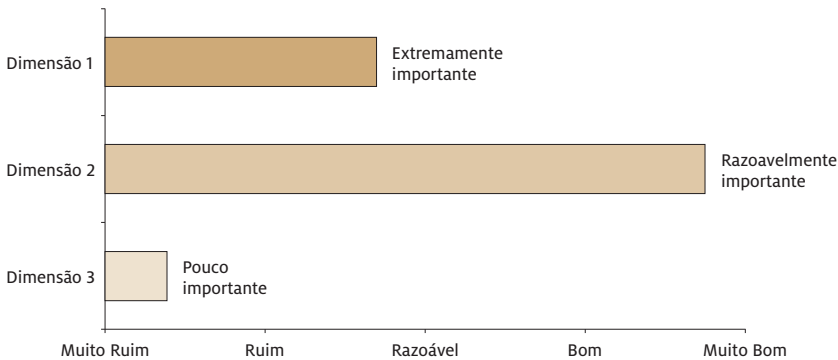
- Excelente: muito superior em relação a todas as evidências coletadas; pode ser considerado como tendo atingido o ponto ideal, ou chegado muito próximo dele.
- Bom: sólido em relação às evidências coletadas, mas com pequenas limitações ou problemas que o distanciam um pouco mais da situação ideal.
- Razoável: inconsistência nas evidências, positivas em sua maioria; atende razoavelmente às expectativas; nenhum dos problemas encontrados é grave.
- Ruim: maioria das evidências traz resultado negativo, apesar de haver poucos resultados positivos; alguns dos problemas encontrados são considerados graves.
- Muito ruim: todas as evidências trazem resultado negativo; vários dos problemas encontrados são considerados graves.

As rubricas avaliativas podem ser tanto genéricas quanto específicas. As genéricas, parecidas com as descrições apresentadas acima, podem ser usadas para definir o desempenho do “avaliando” em todas as dimensões incluídas no estudo. Os avaliadores teriam então que apresentar, detalhadamente, as justificativas que os levaram a determinar que o desempenho do programa em uma determinada dimensão foi “bom” em vez de “ruim” ou “excelente”. As rubricas específicas são aquelas construídas utilizando as evidências pertinentes a cada uma das dimensões analisadas, sendo, portanto, diferentes para cada dimensão. Podem variar também em termos de complexidade, dependendo do número de evidências incluídas. Um exemplo de rubrica avaliativa específica será apresentado na p. 108.

3.3. DETERMINAÇÃO DO NÍVEL DE IMPORTÂNCIA DE CRITÉRIOS, INDICADORES E/OU COMPONENTES

Ao considerar as diferentes dimensões ou componentes de um “avaliando”, fica óbvio que nem todas possuem o mesmo nível de importância. Usando um exemplo extremo, consideremos um programa que tem como principal objeto diminuir a pobreza rural. É provavelmente mais importante que as famílias tenham aumentado de forma sustentável sua capacidade de produzir mais comida e de se alimentar melhor, do que terem passado a participar mais das reuniões da associação comunitária de onde vivem.

Saber que dimensões ou componentes são mais importantes é essencial para determinar que melhorias precisam ser priorizadas, a fim de identificar se os problemas/limitações encontrados são sérios ou menos importantes e/ou determinar se um “avaliando” que tem um desempenho inconsistente (ou seja, com bons e maus resultados) está mais para o lado positivo ou negativo. A figura abaixo traz um exemplo hipotético de representação de desempenho de um programa em relação a três dimensões, cuja importância está explicitada: >



Fonte: adaptado de Davidson (2005; p. 101).

Um desempenho não muito bom na dimensão 3, considerada pouco importante, pode não ser um grande problema. Por outro lado, um desempenho ruim na dimensão 1, considerada extremamente importante, pode significar um desastre para o programa. Se os avaliadores não tivessem determinado o nível de importân-

cia dessas dimensões, os coordenadores ou os financiadores deste programa hipotético poderiam concluir que a dimensão 3 seria a que precisaria de maior atenção imediata, por estar em situação pior em comparação com as outras. Na realidade, a dimensão 1 é a que provavelmente precisaria ser priorizada, tendo em vista sua importância para o programa.

Há diferentes técnicas para determinação do nível de importância de dimensões ou componentes de um “avaliando”. Entre elas destacam-se: (i) votação realizada pelos principais interessados no “avaliando”; (ii) consulta a alguns interessados-chave que conhecem bem o “avaliando”; (iii) uso de evidências da literatura; (iv) opinião de especialistas; (v) uso de evidências coletadas durante estudos de necessidades do público-alvo ou para seleção dos “valores” da avaliação; e (vi) uso de teorias de impacto do programa e de conexões causais. A atribuição de pesos para caracterizar o nível de importância das diferentes dimensões pode ser feita tanto quantitativa quanto qualitativamente. A descrição mais detalhada sobre estas estratégias foge dos objetivos desta publicação e pode ser encontrada em Scriven (1991) ou em Davidson (2005).

3.4. SÍNTESE DE RESULTADOS PARA CHEGAR A CONCLUSÕES AVALIATIVAS

Síntese avaliativa é o processo de combinar uma série de evidências complexas sobre o desempenho ou a situação de um “avaliando” em diferentes dimensões ou componentes para se chegar a conclusões avaliativas. As rubricas avaliativas são consideradas também como sínteses avaliativas. Porém, há casos em que as evidências são muito diversificadas – provêm de diferentes fontes de informação ou são de múltiplas naturezas (quantitativas e qualitativas) –, tornando-se complexas demais para que se possam elaborar rubricas avaliativas capazes de abranger de forma correta todos os aspectos importantes envolvidos.

Há várias técnicas para se fazer sínteses avaliativas. Antes de escolher uma delas é preciso saber se o objetivo é gerar conclusões avaliativas sobre um único “avaliando” ou se a intenção é comparar o desempenho de diferentes “avaliandos”. Se a síntese destina-

-se a determinar o mérito de um único “avaliando”, os tipos mais comuns de perguntas que podem ser respondidas são: “Quão bom foi o desempenho do ‘avaliando’ no geral?”; “Quão bom foi o desempenho do ‘avaliando’ em relação às principais dimensões de interesse?”; “Valeu a pena implementar este componente do ‘avaliando’, levando em conta o investimento feito?”. Exemplos de sínteses que visam classificar diferentes “avaliandos” podem incluir o ranqueamento de universidades, comparações entre diferentes programas para reduzir a pobreza rural, seleção de candidatos para uma vaga de emprego, ou comparação entre diferentes produtos.

Existem as sínteses quantitativas e as qualitativas. As quantitativas usam pesos numéricos tanto para determinar a importância das diferentes dimensões (por exemplo: peso 1 para dimensões pouco importantes e peso 4 para as extremamente importantes), quanto para definir a pontuação do desempenho do “avaliando” (por exemplo: 1 ponto para desempenho muito ruim e 5 pontos para desempenho excelente). Nesses casos, as conclusões avaliativas são determinadas a partir de multiplicações de pesos e pontuações, além de somatórios e médias dos resultados para cada dimensão e/ou “avaliando”. As qualitativas usam símbolos ou categorias para representar o nível de importância (peso) de cada dimensão, bem como do desempenho do “avaliando”. Chega-se às conclusões avaliativas através de um processo complexo de reflexões lógicas e não pelo resultado de multiplicações e médias numéricas.

Quando se tem um número razoavelmente grande de dimensões ou aspectos compondo as avaliações, os métodos qualitativos são melhores do que os quantitativos. Nessas situações, as sínteses numéricas podem levar a conclusões inválidas, pois tendem a diluir a contribuição das dimensões realmente importantes no resultado final e acabam deixando que o bom desempenho em várias dimensões pouco importantes influencie mais o resultado final do que o mau desempenho em dimensões realmente importantes.

Outro elemento relevante nas sínteses diz respeito à criação de “barreiras”. Elas constituem o desempenho mínimo que um “avaliando” deve ter em determinadas dimensões ou aspectos estratégicos para ser considerado aceitável. Determinam também que se um “avaliando” não tiver um desempenho pelo menos razoável em uma

dimensão-chave, será considerado, como um todo, ruim, não importando se teve ótimo desempenho em outras dimensões. Usando um exemplo extremo: no processo seletivo de um profissional, se ele não apresentar postura ética adequada, não deve ser contratado, mesmo que seja considerado tecnicamente excelente. Pelas consequências importantes que as “barreiras” têm para um processo avaliativo, elas devem ser usadas com muito cuidado e parcimônia.

Mais detalhes sobre as diferentes metodologias para a realização de sínteses podem ser encontrados nos livros de Scriven (1991) e de Davidson (2005).

Os quatro métodos citados anteriormente são exemplo das principais metodologias específicas para a avaliação – aquelas que efetivamente contribuem para determinar a qualidade, a importância ou o valor de determinado “avaliando”. Não podem ser considerados como pertencentes à mesma categoria os métodos estatísticos ou os outros métodos padrão de pesquisa, como entrevistas, observações, questionários, análise de conteúdo ou mesmo de inferência causal. Certamente os avaliadores fazem uso constante de tais métodos, mas eles não são especificamente avaliativos, porque não tratam da questão dos “valores” de maneira explícita. Ou seja, os profissionais podem recorrer a eles (sejam quantitativos ou qualitativos), mas não estarão, necessariamente, fazendo avaliações. Entretanto, se empregarem uma ou mais das quatro metodologias descritas acima, certamente estarão realizando avaliação.

Em síntese: as metodologias específicas da avaliação produzem, de forma sistemática e transparente, evidências sólidas sobre o desempenho de um “avaliando” e permitem chegar a conclusões avaliativas com base nos “valores” selecionados. Como produzir, na prática, avaliações de qualidade, que respondam diretamente às perguntas avaliativas formuladas e que gerem ação?

4. PONTOS-CHAVE PARA SE FAZER AVALIAÇÕES VÁLIDAS, RELEVANTES E ÚTEIS

As ideias apresentadas neste texto estão muito influenciadas pelo trabalho de Michael Scriven e de E. Jane Davidson. Scriven pode ser considerado como um dos pais da avaliação, tendo sido res-

ponsável por estabelecer boa parte da fundamentação teórica e prática do campo. Seu livro *Evaluation thesaurus*¹ (Scriven, 1991) é, até hoje, a principal referência sobre lógica e metodologia da avaliação. Outra referência fundamental de sua autoria, o *Key evaluation checklist*² (Scriven, 2011), apresenta em detalhe os aspectos-chave que precisam ser considerados em qualquer avaliação, tendo sido Jane Davidson quem melhor conseguiu traduzir tais construções para a prática. Ao longo dos últimos anos, ela vem estabelecendo as bases da metodologia batizada de “avaliações relevantes” (Davidson, 2012).

A combinação do *Key evaluation checklist* e da metodologia de avaliações relevantes é o que orienta os pontos-chave para realizar boas avaliações apresentados a seguir. Sabe-se que uma avaliação é válida, relevante e útil quando:

- é claramente focada nas mudanças que o “avaliando” está tentando produzir ou alcançar;
- vai direto ao cerne do que é realmente importante no “avaliando”, em vez de se perder em aspectos menos importantes;
- gera informações úteis nos níveis estratégico e no prático;
- influencia e esclarece pensamentos e decisões; e
- traz *insights* que ajudam na tomada de decisões sobre quais ações devem ser tomadas.

Em contraste, avaliações de baixa relevância e utilidade tendem a:

- ir direto para o planejamento dos métodos de pesquisa para coletar e analisar dados, sem formar uma ideia clara do que é e para quem servirá a avaliação;
- abordar de forma incorreta o envolvimento de diversos grupos interessados ou afetados diretamente pela avaliação: às

1. “Tesouro da avaliação”, em tradução livre.

2. “Lista de checagem-chave da avaliação”, em tradução livre.

vezes são envolvidos muitos, poucos ou as pessoas erradas, ou no momento inadequado;

- formular perguntas muito específicas e descritivas, sem abordar os pontos mais importantes que fazem o “avaliando” ser bom, razoável ou ruim;
- apresentar muitos dados quantitativos e/ou qualitativos meramente descritivos, sem concluir se a situação ou o desempenho do “avaliando” foi boa/bom; e
- deixar o leitor incerto sobre o que precisa ser feito em relação ao “avaliando” – pontos positivos e negativos podem ter sido identificados, mas não está claro quão importantes ou sérios são, ou não há pistas claras sobre o que está funcionando melhor ou pior para quem, onde ou por quê.

São seis os pontos-chave que devem ser levados em consideração para se ter certeza de que se obterá uma avaliação de alta qualidade. Para ilustrar alguns deles, vejamos exemplos práticos de um projeto em nível nacional, apoiado pela Fundação Roberto Marinho (mais informações sobre esta avaliação estão presentes nas pp. 117 a 124).

1. O primeiro ponto-chave é *entender claramente quais os principais propósitos da avaliação* para as pessoas que têm interesse direto nos seus resultados – os *stakeholders*. É necessário que os avaliadores identifiquem quem precisa saber o que, com qual objetivo e quando a informação será mais útil. Para fazer isso de maneira adequada, precisam interagir com os *stakeholders*, ajudando-os a identificar se a avaliação está focada mais em gerar informações que sirvam para melhorar o “avaliando” ou para tomar decisões sobre o futuro (por exemplo: expandir, reduzir, manter etc.) – ou ambos. Um exercício importante para ajudar nesse processo é reunir os principais *stakeholders* e solicitar que reflitam conjuntamente e que, a seguir, explicitem as discussões em uma matriz, como no exemplo a seguir, para esclarecer os propósitos da avaliação. >

QUEM precisa aprender?	O QUE precisa ser aprendido?	Com que PROPÓSITO?	QUANDO a informação será útil?
Gestores da fundação	<ul style="list-style-type: none"> • Relevância dos resultados do projeto para o público-alvo 	<ul style="list-style-type: none"> • Prestar conta aos parceiros mantenedores • Aprimorar os modelos de intervenção • Prospectar novos parceiros 	<ul style="list-style-type: none"> • Reuniões anuais do Conselho (dezembro) • Prestação de contas aos financiadores (semestralmente)
Equipe coordenadora/ implementadora	<ul style="list-style-type: none"> • Metodologias empregadas são eficazes e eficientes • Relevância dos resultados do projeto para o público-alvo • Lacunas de conteúdo no projeto • Em que medida o projeto fortalece as organizações parceiras 	<ul style="list-style-type: none"> • Melhorar o projeto ao longo de sua implementação • Melhorar futuros projetos • Gerar ideias para novos conteúdos e projetos • Contribuir para fortalecer as organizações parceiras 	<ul style="list-style-type: none"> • Reuniões semestrais da equipe do projeto • Planejamento de novo projeto (dezembro 2013)
Organizações parceiras	<ul style="list-style-type: none"> • Em que medida o projeto fortalece as organizações parceiras • Relevância dos resultados do projeto para os educadores 	<ul style="list-style-type: none"> • Melhorar futuros projetos • Gerar ideias para novos conteúdos e projetos 	<ul style="list-style-type: none"> • Ao longo de 2013/2014

Fonte: o autor, com autorização da Fundação Roberto Marinho.

2. O segundo elemento-chave requer o planejamento e a adoção de estratégias para *envolver os principais stakeholders na avaliação*: quem precisa estar envolvido, quando e como? É necessário não só identificar quem são as pessoas, mas também criar mecanismos para viabilizar sua participação. É importante aqui pensar no envolvimento estratégico e inteligente, que alcance o equilíbrio com a viabilidade da avaliação dentro das limitações de tempo e de outros recursos – restrições presentes em qualquer processo avaliativo. O quadro a seguir apresenta o mapeamento dos principais *stakeholders* e a definição de quem deve ser envolvido em que fases da avaliação, bem como de que forma se dará o envolvimento. >

Em que ETAPAS envolver?	QUEM envolver?*	COMO envolver?
Perguntas avaliativas	(1); (2); (3); (4); (5)	(1) e (2) reuniões presenciais; (3) consulta à distância (Skype/telefone/e-mail); (4) e (5) consulta (reunião de apresentação e discussão do documento final).
Definição de qualidade e valor (rubricas)	(1); (2); (4); (7)	(1), (2) e (7) reuniões presenciais; (4) consulta (reunião de apresentação e discussão do documento final).
Coleta de evidências	(1); (2); (3)	(1), (2) e (3) apoio logístico.
Interpretação e discussão de resultados	(1); (2); (3); (4); (5); (6)	Reuniões presenciais para discussão dos resultados da avaliação.
Planejamento para ação	(1); (2); (3); (4)	Reuniões presenciais para discutir o uso dos resultados da avaliação.
* (1) Equipe gestora da fundação; (2) Equipe gestora da área onde está localizado o projeto; (3) Equipe implementadora do projeto; (4) Organizações parceiras; (5) Conselheiros da fundação; (6) Parceiros mantenedores; (7) Especialistas na área do projeto		

Fonte: o autor, com autorização da Fundação Roberto Marinho.

- O terceiro elemento-chave é *definir as perguntas estratégicas que irão orientar todo o processo avaliativo*. Devem ser perguntas que tratem da qualidade e do valor do programa e não apenas descrevam o que aconteceu. Percebam a diferença entre indagar “Que mudanças o programa conseguiu provocar?” e “Quão importantes foram as mudanças provocadas pelo programa?”. A primeira formulação é meramente descritiva e não gera informações que possibilitarão chegar a conclusões sobre quão bom e relevante o programa é. Não há dúvida de que, em qualquer processo avaliativo, será necessário saber quais as mudanças provocadas pelo programa. Porém, se não se investiga a importância das mesmas, torna-se difícil saber se foi um sucesso retumbante ou um total fracasso, ou algo entre esses extremos. Consequentemente, a tomada de decisão sobre o que fazer com o programa (por exemplo: melhorar, expandir, terminar etc.) será pouco fundamentada.

No quadro a seguir, apresentamos alguns exemplos de perguntas avaliativas para a avaliação de um programa na área de educação e mobilização social: >

<p>1. Em que medida o projeto contribui para atender a necessidades e prioridades das organizações parceiras e dos potenciais beneficiários finais (educadores, crianças, famílias e comunidade)?</p>
<p>2. Quanto bem desenhado e implementado foi o projeto?</p> <p>2.1. As estratégias e critérios adotados para selecionar as organizações parceiras foram os mais acertados para identificar as organizações mais efetivas onde implementá-lo?</p> <p>2.2. O apoio inicial e ao longo do projeto oferecido pela equipe gestora às organizações parceiras foi suficiente e efetivo para garantir o alcance de resultados significativos?</p> <p>2.3. O apoio inicial e ao longo do projeto oferecido pelas organizações parceiras às outras organizações participantes foi efetivo e suficiente?</p> <p>2.4. Os mecanismos de monitoramento presencial e a distância adotados e o uso dos resultados desse acompanhamento foram adequados para a boa execução do projeto?</p>
<p>3. Qual a qualidade e importância dos resultados alcançados pelo projeto junto aos educadores, crianças, famílias/comunidade e organizações parceiras?</p> <p>3.1. Quanto importante foi sua contribuição para aumentar o conhecimento e melhorar o trabalho dos educadores nas escolas e em outros espaços educativos?</p> <p>3.2. Em que medida ajudou a fortalecer as organizações parceiras?</p>
<p>4. Quanto bem o projeto funcionou nas diferentes realidades geográficas (por exemplo: zona urbana x rural; região Nordeste x região Sudeste) e para os diferentes tipos de organizações parceiras (por exemplo: ONGs x secretarias municipais)? Por quê?</p>
<p>5. Os resultados do projeto justificaram os custos incorridos com seu planejamento e implementação e com o tempo e outros recursos investidos pela equipe de gestão da fundação e das organizações parceiras para sensibilização, apoio e monitoramento?</p>
<p>6. Qual a probabilidade de as mudanças positivas provocadas pelo projeto perdurarem após a finalização das ações planejadas?</p>

Fonte: o autor, com autorização da Fundação Roberto Marinho.

4. O quarto elemento envolve *obter uma combinação acertada de evidências (quantitativas e qualitativas) capazes de dar respostas convincentes às perguntas avaliativas elaboradas*. É essencial estabelecer os parâmetros para definir “quanto bom é bom”. Ou seja, que critérios deverão ser usados para concluir se o desempenho do programa em relação às perguntas avaliativas foi excelente, bom, razoável, ruim ou péssimo/inaceitável.

Como mencionado anteriormente, as rubricas avaliativas são um instrumento muito útil para dar conta desse trabalho. Uma estratégia para desenvolvê-las de maneira correta envolve, primeiro, identificar os critérios de mérito do “avaliando”, ou seja, aqueles que definem a situação ideal/desejada para a pergunta avaliativa. Depois, faz-se necessário descobrir quais

evidências precisarão ser coletadas para determinar se os critérios de mérito foram bem cumpridos ou não. O quadro a seguir apresenta um exemplo de critério de mérito e de evidências a serem coletadas. >

Pergunta avaliativa	Critério de mérito	Evidências
2.4. Os mecanismos de monitoramento presencial e a distância adotados e o uso dos resultados desse acompanhamento foram adequados para a boa execução do projeto?	<ul style="list-style-type: none"> Os mecanismos de monitoramento adotados foram efetivos para informar, no momento certo, sobre a necessidade de ajustes na implementação de ações que contribuíram para resultados significativos do projeto. 	<ul style="list-style-type: none"> Parecer da equipe de avaliação externa sobre a qualidade das estratégias de monitoramento Percepção da equipe de coordenação da fundação sobre a qualidade, eficiência e usabilidade das estratégias de monitoramento do projeto Percepção da equipe de implementação sobre a qualidade, eficiência e usabilidade das estratégias de monitoramento Percepção dos representantes das organizações parceiras sobre a qualidade, eficiência e usabilidade das estratégias de monitoramento Exemplos de como os resultados dos relatórios de monitoramento escritos (por exemplo: relatório formal impresso, comunicação por <i>e-mail</i> etc.) e orais (por exemplo: apresentações formais, conversas por telefone etc.) foram utilizados para melhorar a implementação do projeto

Fonte: o autor, com autorização da Fundação Roberto Marinho.

As rubricas avaliativas são então elaboradas considerando a melhor combinação das diferentes evidências identificadas para determinar o que seria um desempenho excelente, bom, razoável, ruim ou péssimo/inaceitável. O quadro a seguir traz um exemplo prático de rubrica avaliativa elaborada para responder à pergunta avaliativa 2.4. do exemplo, utilizando as evidências incluídas no quadro anterior. >

Conclusão	Rubricas
Excelente	<ul style="list-style-type: none"> • Os relatórios de monitoramento são considerados de alta qualidade (formato e conteúdo) pelos avaliadores externos, sem problemas ou omissões importantes. • Os relatórios foram socializados com todos da equipe de implementação do projeto e com quase todas (95% ou mais) as organizações parceiras. • Há um número muito grande de exemplos de uso dos resultados do monitoramento feito pela equipe de implementação, bem como pelas organizações parceiras para melhorar a implementação do projeto.
Bom	<ul style="list-style-type: none"> • Os relatórios de monitoramento são considerados de boa qualidade (formato e conteúdo) pelos avaliadores externos, com quase nenhum problema ou omissões importantes. • Os relatórios foram socializados com todos da equipe de implementação do projeto e com um grande número (80% a 95%) de organizações parceiras. • Há um número grande de exemplos de uso dos resultados do monitoramento feito pela equipe de implementação, bem como pelas organizações parceiras, para melhorar a implementação do projeto.
Razoável	<ul style="list-style-type: none"> • Os relatórios de monitoramento são considerados de qualidade razoável (formato e conteúdo) pelos avaliadores externos, mas existem alguns problemas e/ou omissões importantes. • Os relatórios foram socializados com todos da equipe de implementação do projeto e com um número razoável (65% a 80%) de organizações parceiras. • Há um número razoável de exemplos de uso dos resultados do monitoramento feito pela equipe de implementação, bem como pelas organizações parceiras, para melhorar a implementação do projeto.
Ruim	<ul style="list-style-type: none"> • Os relatórios de monitoramento são considerados de baixa qualidade (formato e conteúdo) pelos avaliadores externos e contêm diversos problemas e/ou omissões importantes. • Os relatórios não foram socializados com todos da equipe de implementação do projeto e um número pequeno (menos que 65%) de organizações parceiras tiveram acesso a eles. • Há poucos exemplos de uso dos resultados do monitoramento feito pela equipe de implementação, bem como pelas organizações parceiras, para melhorar a implementação do projeto.

Fonte: o autor, com autorização da Fundação Roberto Marinho.

As rubricas podem variar no número de categorias e no nível de complexidade, dependendo da quantidade e complexidade das evidências. Há rubricas genéricas que podem ser utilizadas como base para determinar o nível de desempenho do “avaliando” em relação a todas as perguntas avaliativas ou a dimensões de impacto selecionadas. Um exemplo de rubrica avaliativa genérica pode ser encontrado no artigo de Davidson (2012; p. 27). No processo de construção das rubri-

cas, é essencial o envolvimento dos principais *stakeholders* do “avaliando”. Só assim elas se tornarão precisas, possibilitando também uma oportunidade para que os responsáveis e usuários do “avaliando” formem uma ideia mais clara sobre o que efetivamente estão fazendo e que mudanças visam alcançar.

5. O quinto elemento envolve *elaborar relatórios claros e sucintos dos resultados da avaliação*. Relatórios descritivos e longos, inspirados nos relatórios acadêmicos, em geral, não são muito úteis. Como produzir um relatório que vá direto ao ponto, sem se perder em detalhes triviais, é um desafio que os avaliadores precisam encarar. Um bom relatório de avaliação caracteriza-se principalmente por responder de forma direta e sintética às perguntas avaliativas, usando a linguagem que os principais usuários da avaliação entendem, com diferentes “camadas” que interessem aos diferentes públicos leitores.

Para isso, o relatório precisa ter:

- (i) um sumário executivo entre uma e três páginas, respondendo resumidamente às perguntas avaliativas ou discutindo o desempenho do “avaliando” em cada uma de suas dimensões ou componentes, e trazendo as principais conclusões e ideias para subsidiar a tomada de decisões estratégicas;
- (ii) um corpo principal contendo não mais que 30 páginas, focado nas respostas às perguntas avaliativas e na apresentação das evidências que basearam as conclusões avaliativas, incluindo representações gráficas que facilitem o entendimento do material; e
- (iii) anexos contendo todos os detalhes metodológicos, incluindo as etapas percorridas para elaborar as perguntas avaliativas, identificar e selecionar as evidências e rubricas, além dos métodos e instrumentos utilizados para coletar e analisar dados.

6. Por fim, a avaliação deve ser capaz de *gerar ideias e insights que ajudem os stakeholders a agir*. Na grande maioria dos relatórios, os avaliadores devem oferecer recomendações em nível micro, que não incorram em custos adicionais significativos para o programa. Devem deixar as recomendações macro apenas para os casos em que a equipe de avaliação tiver conhecimento profundo não só da parte técnica do “avaliando”, mas também do contexto político e financeiro das organizações e dos outros atores envolvidos – o que é muito difícil de se conseguir.

Mais detalhes sobre os seis pontos-chave de uma avaliação de qualidade podem ser encontrados no artigo publicado por Davidson (2012).

5. EXEMPLOS PRÁTICOS

Há três experiências recentes nas quais o autor teve a oportunidade de aplicar os conceitos aqui descritos. A primeira foi uma série de avaliações de projetos apoiados pela Heifer International, uma ONG baseada nos Estados Unidos que atua em mais de 40 países, apoiando organizações locais que desenvolvem projetos de redução de pobreza e de proteção do meio ambiente junto a pequenos agricultores.

O modelo básico de intervenção inclui a doação, a pequenos agricultores, de animais de grande, médio e pequeno portes (por exemplo, respectivamente: vacas e camelos; cabras e porcos; e abelhas e minhocas); capacitação para o manejo dos animais; e atividades de mobilização, como a criação de organizações comunitárias. Um aspecto marcante do modelo da Heifer é o compromisso assumido pelas pessoas que recebem os animais de repassar o mesmo número, com a mesma qualidade, a outras pessoas, em determinado período de tempo.

A avaliação foi implementada pelo Centro de Avaliação da Universidade de Western Michigan entre 2007 e 2011 e englobou quase 200 projetos em 23 países, envolvendo mais de 90 profissionais, entre avaliadores externos, consultores locais e tradutores³.

3. As avaliações foram dirigidas por Michael Scriven e gerenciadas pelo autor deste artigo.

Foi um processo muito complexo, que, na essência, definiu, junto com as equipes da Heifer no nível central e nos diferentes países, os principais “valores” (seis no total) de todo o trabalho realizado pela organização e por seus parceiros.

A partir daí, foram criados mais de cem critérios e indicadores que explicitavam tais “valores” em cada projeto. Foram criadas rubricas avaliativas que possibilitaram a produção de conclusões avaliativas sobre cada indicador, critério e grupo de valores, baseadas em evidências quantitativas e qualitativas levantadas em cada projeto. Houve o cuidado de realizar reuniões para a discussão dos resultados das avaliações com os coordenadores dos projetos em cada país e com o escritório central nos Estados Unidos. Também foram desenvolvidas estratégias para representar graficamente os resultados e assim facilitar sua comunicação e uso.

No Brasil, dois processos avaliativos representam fielmente a aplicação da abordagem descrita acima. O primeiro já dura quase cinco anos e está sendo realizado junto ao Programa Primeiríssima Infância da Fundação Maria Cecília Souto Vidigal no Estado de São Paulo (o estudo de caso encontra-se nas pp. 125 a 134). O programa tem como objetivo principal mobilizar a sociedade e capacitar profissionais que atuam junto a gestantes, crianças de 0 a 3 anos e suas famílias em temas-chave dessa etapa da vida, buscando assim contribuir para que novas gerações tenham o apoio necessário para seu pleno desenvolvimento. A implementação se dá em parceria com diversos municípios e atualmente conta também com a parceria da Secretaria de Estado de Saúde. A avaliação utilizou diversas estratégias, cabendo destacar:

- desenvolvimento de um modelo de impacto, em consulta com a equipe de gerentes da Fundação Maria Cecília Souto Vidigal, especialistas e equipes dos projetos locais – o modelo definiu as oito principais dimensões nas quais o programa deveria produzir mudanças em prol da primeira infância;
- criação de 50 indicadores diretamente conectados com a qualidade dessas dimensões;
- criação de rubricas avaliativas combinando evidências quantitativas e qualitativas para cada indicador; e

- elaboração de relatórios preliminares para cada município, discutidos em reuniões devolutivas em que se confirmam os achados da avaliação e se promove o debate entre as equipes locais e da fundação, para que sejam tomadas decisões visando aperfeiçoar o trabalho.

Muitos outros aspectos interessantes desse processo estão sendo sintetizados em uma publicação que deverá estar disponível para o público em breve.

O exemplo mais recente de utilização da mesma abordagem é o da avaliação do Projeto Maleta Infância, do Canal Futura, da Fundação Roberto Marinho (o estudo de caso encontra-se nas pp. 117 a 124). Esta é, provavelmente, a experiência em que se tem conseguido aplicar de forma mais didática os seis pontos-chave aqui descritos.

Esse projeto tem âmbito nacional e está focado na promoção do debate e de ações por diferentes organizações que atuam no campo da Educação de crianças de 0 a 11 anos. A estratégia principal é o envio de uma maleta com vários materiais audiovisuais sobre Educação Integral, principalmente programas do Canal Futura, para aproximadamente 550 organizações em todo o País. Elas recebem apoio de mobilizadores comunitários do Canal Futura, que promovem a troca de experiências entre organizações para aprimoramento e convergência do trabalho local.

A avaliação teve início em janeiro de 2013 e, até o momento, foi realizado um processo participativo para identificação dos interessados (estratégias de envolvimento), formulação de perguntas avaliativas, identificação de critérios e evidências, e elaboração de rubricas avaliativas. Há outro componente interessante no processo, que é o estudo de casos de sucesso, visando analisar determinado número de organizações que conseguiram realizar o trabalho com êxito e outras que enfrentaram dificuldades para alcançar resultados positivos. A informação gerada por esse estudo ajudará a fundação a entender melhor para quem, onde e por que o projeto funcionou melhor, para orientar edições futuras do projeto. O processo está em implementação e terá duração estimada de dois anos.

6. PRINCIPAIS DESAFIOS

Na prática, os desafios contextuais são sempre os mais difíceis para implementação desta abordagem avaliativa pelos profissionais. A situação mais complicada ocorre quando há diferentes perspectivas ou mesmo conflitos entre os *stakeholders*. Na maioria dos casos, é possível superar as diferenças por meio do diálogo e da negociação, estabelecendo acordos para garantir que a avaliação seja realizada com a qualidade necessária. Há, entretanto, casos mais sérios, felizmente a minoria, nos quais as divergências são tão extremas que chegam a comprometer a viabilidade da avaliação – seja pelo boicote ao acesso a informações cruciais, seja pela criação de movimento para desacreditar o processo e criar barreiras para que seus resultados sejam utilizados.

Outras situações desafiadoras envolvem tensões em relação a limitações orçamentárias, ausência de profissionais capacitados para realizar avaliações dessa natureza e pouco tempo para implementar de forma adequada todas as etapas necessárias, principalmente as que exigem o envolvimento dos interessados no estudo.

7. COMENTÁRIOS FINAIS

A avaliação é um trabalho complexo, que tem papel fundamental para a melhoria de processos e de resultados em todos os campos do conhecimento. Entretanto, só conseguirá cumprir esse potencial de maneira plena, se for realizada corretamente. Tanto os avaliadores quanto os *stakeholders* precisam estar conscientes sobre os aspectos que influenciam diretamente a qualidade das avaliações. Precisam também se livrar de ideias equivocadas que têm influenciado negativamente o campo, principalmente as herdadas da academia. Vale lembrar as três mensagens básicas enfocadas neste artigo:

1. A avaliação não se restringe a medir indicadores. Ela visa essencialmente apresentar respostas bem fundamentadas para perguntas importantes e realmente avaliativas.

2. Há metodologias específicas do campo da avaliação que ajudam a navegar pelos meandros complexos dos “avaliandos”, principalmente os programas sociais. São elas que tornam um estudo realmente avaliativo.
3. Para que uma avaliação seja válida, relevante e útil é preciso:
 - ter clareza sobre o que se quer avaliar;
 - identificar as pessoas-chave interessadas na avaliação, criando estratégias para envolvê-las nos momentos estratégicos;
 - formular perguntas realmente avaliativas, que tratem da essência do que o “avaliando” se propõe a realizar;
 - estabelecer meios concretos para responder clara e diretamente às perguntas avaliativas;
 - criar relatórios claros e sucintos que respondam diretamente às perguntas importantes; e
 - gerar ideias que ajudem as organizações a tomarem boas decisões.

BIBLIOGRAFIA

- DAVIDSON, E. J. **The nuts and bolts of sound evaluation**. Thousand Oaks, CA: Sage, 2005.
- _____. Tornar as avaliações estrategicamente práticas e relevantes. In: FUNDAÇÃO ITAÚ SOCIAL, FUNDAÇÃO ROBERTO MARINHO E MOVE (Orgs.). **A relevância da avaliação para o investimento social privado**. São Paulo: Fundação Santillana, 2012.
- SCRIVEN, M. **Evaluation thesaurus**. Newbury Park, CA: Sage, 1991.
- _____. **The key evaluation checklist**, abr. 2011. Disponível em: <http://michaelscriven.info/images/KEC_4.18.2011.pdf> (último acesso em setembro de 2013).

O uso da abordagem participativa na avaliação do Projeto Maleta Infância

FUNDAÇÃO ROBERTO MARINHO – CANAL FUTURA

1. O CANAL FUTURA

O Futura é um projeto social de comunicação e educação, de interesse público, gerado e construído por parceiros da iniciativa privada e do terceiro setor. É comprometido com a transformação social, veiculando na tela da TV conteúdos educativos. Para além da programação televisiva, o Futura atua nos territórios de forma presencial, através da área denominada *Mobilização e Articulação Comunitária*, que conta com educadores sociais (mobilizadores) espalhados pelo Brasil.

Essa equipe está em constante articulação com instituições públicas, privadas e não governamentais, mediando um diálogo crítico e contínuo, que permite a disseminação e o uso do conteúdo do canal para além da tela da TV e a inclusão na programação de temas e abordagens na perspectiva de diferentes grupos sociais.

2. A PESQUISA E A AVALIAÇÃO NO CANAL FUTURA

A avaliação percorre a história do Canal Futura, que realiza continuamente acompanhamentos, pesquisas e avaliações quantitativas e qualitativas de suas iniciativas.

São realizadas pesquisas anuais sobre o perfil do público e a audiência, grupos focais sobre programas e avaliações constantes de projetos especiais. Essa sistemática gera um círculo virtuoso para o canal, que planeja suas atividades utilizando as informações das pesquisas e avaliações que realiza.

As escolhas avaliativas – como o propósito, as perguntas, os métodos etc. – são fruto de diálogo que conecta pessoas, ideias, redes e instituições. Podemos dizer, nesse sentido, que a abordagem das avaliações do Canal Futura está atrelada a sua própria história, aos *stakeholders* e a seus valores.

3. O PROJETO MALETA INFÂNCIA

A Educação Integral se estabelece como prioritária e orienta tanto o desenho de programas federais, como o *Mais Educação* e o *Plano de Desenvolvimento da Educação*¹, quanto ações do investimento social privado. Entretanto, apesar dos esforços, persistem no Brasil diversos desafios para a sua implementação, entre eles a carência de materiais apropriados e o desafio de aprofundar a compreensão dos pressupostos relacionados ao conceito de Educação Integral.

Ao longo de seus quinze anos no ar, o Canal Futura inseriu-se no debate e desenvolveu ações para contribuir com a qualificação da educação por meio de produções audiovisuais e da inserção, em sua grade de programação, de temáticas ligadas às áreas de educação ambiental, lazer, direitos humanos, ciências, linguagens, artes e saúde, entre outras.

Nesse contexto, e em parceria com organizações da sociedade civil, o Canal Futura estabeleceu um programa denominado Projeto Maleta, já em sua quinta edição em 2013.

A versão atual, denominada Maleta Infância, tem o intuito de estimular o debate e ações educativas que favoreçam a Educação Integral da criança. A Maleta Infância é voltada a educadores e cuidadores, tanto de escolas formais como de espaços educacio-

1. Programas lançados com o intuito de otimizar as ações e os investimentos já existentes no País para que complementem a formação escolar com uma visão integradora do ensino.

nais alternativos que existem nas comunidades. Esses educadores e cuidadores podem atuar diretamente com crianças, bem como com outros educadores, cuidadores, comunidades escolares e organizações que trabalhem com a implementação de políticas públicas relacionadas à Educação Integral e à família.

A definição dos conteúdos das edições do Projeto Maleta é feita por meio de fóruns temáticos realizados com representantes de ONGs, instituições sociais e profissionais que atuam nas áreas prioritizadas em cada projeto. Nesses fóruns são estabelecidos os pontos essenciais a serem abordados nos conteúdos dos produtos audiovisuais incluídos em cada Maleta. São reunidas na Maleta tanto produções do Canal Futura quanto materiais das instituições parceiras que são referências na temática. Quando necessário, novos materiais são especificamente produzidos.

Por meio da equipe de mobilização comunitária do Canal Futura, o projeto estabelece parcerias com organizações de referência em várias regiões do Brasil. No caso do Projeto Maleta Infância, são 50 organizações parceiras entre ONGs, universidades e outras instituições que trabalham com educadores, creches e Ensino Básico, bem como organizações formadoras que implementam políticas públicas de Ensino Integral. Os critérios para a seleção desses parceiros são: (i) instituições que têm atuação em rede, com capilaridade de ações em um ou mais territórios; (ii) organizações de referência na temática; e (iii) instituições que atuam estrategicamente junto às redes de Educação Básica.

Cada uma das 50 organizações de referência selecionadas tem como responsabilidade identificar cerca de dez instituições locais que trabalhem direta ou indiretamente com ações de Educação Integral para a infância. As organizações de referência recebem uma Maleta Completa, enquanto as demais instituições recebem Maletas Básicas que serão usadas em ações pedagógicas. Ao aderir ao projeto, as instituições firmam um termo de cooperação técnica com o Canal Futura e passam a contar, durante um ano, com o apoio dos mobilizadores comunitários do canal na implementação das maletas por meio de diferentes estratégias de acompanhamento, tanto presenciais quanto virtuais.

4. A AVALIAÇÃO DO PROJETO MALETA INFÂNCIA

A Fundação Roberto Marinho (FRM) realizou, antes da atual edição do Projeto Maleta, estudo exploratório dos processos de implementação e dos resultados alcançados pelas quatro edições anteriores, que trataram dos temas “Beleza”, “Democracia”, “Meio Ambiente” e “Saúde”. O estudo envolveu consultas a uma amostra regional de organizações de referência. Os indicadores adotados pelo estudo permitiram entender dimensões relacionadas ao uso e à finalidade dos materiais, à importância da presença do mobilizador comunitário para a realização das ações do projeto, e à relevância do material para a organização parceira. No entanto, os indicadores não foram suficientemente sensíveis na identificação das contribuições do projeto para possíveis mudanças produzidas.

Decidiu-se então pela construção de um novo modelo de avaliação orientado por perguntas avaliativas estratégicas que levassem à ação, formuladas com base no envolvimento amplo e inteligente dos principais interessados no processo e nos resultados da avaliação. O novo modelo, inspirado no trabalho de E. Jane Davidson², vem sendo aplicado junto ao Projeto Maleta Infância desde o início de 2013.

O primeiro passo do novo modelo incluiu a definição dos principais propósitos da avaliação. Foi feito um mapeamento dos interessados estratégicos pela avaliação e foram definidos: (i) quais as necessidades de informação desses grupos; (ii) por que essas informações são importantes; e (iii) quando tais dados precisam estar disponíveis. O passo seguinte envolveu a identificação dos momentos estratégicos em que membros dos diferentes grupos de interessados na avaliação, explicitados na lista a seguir, precisam ser envolvidos no processo avaliativo.

2. Ver o livro *Evaluation methodology basics* e o artigo da autora “Tornar as avaliações estrategicamente práticas e relevantes”, publicado em *A relevância da avaliação para o investimento social privado* (Fundação Santillana, São Paulo, 2012), disponível para download em pdf em: <<http://www.fundacaosantillana.com.br/images/Avaliação%20investimento.pdf?fileId=8A8A8A833ABBDD9B013ABCBA40C517A1>> e como livro eletrônico em: <http://issuu.com/ed_moderna/docs/avalia__o_investimento?e=2064801/1086043> (últimos acessos em setembro de 2013).

1. Equipe de Desenvolvimento Institucional da Fundação Roberto Marinho (FRM)
2. Equipe de Mobilização Comunitária do Canal Futura/FRM
3. Organizações parceiras (Maleta Infância)
4. Equipe gestora do Canal Futura
5. Secretaria- geral da FRM
6. Parceiros mantenedores
7. Consultores especialistas

Como se visualiza na tabela a seguir, a intensidade e a forma de envolvimento destes atores variam, havendo interações dos mesmos desde a formulação das perguntas avaliativas até a discussão e uso dos resultados. >

FORMAS DE ENVOLVIMENTO DOS PRINCIPAIS INTERESSADOS NA AVALIAÇÃO		
Etapas	Quem envolver?	Como envolver?
Perguntas avaliativas	(1); (2); (3); (4); (5)	(1) e (2) reuniões presenciais; (3) consulta à distância (Skype/telefone/ <i>e-mail</i>); (4) e (5) consulta (reunião de apresentação e discussão do documento final).
Definição de qualidade e valor (rubricas)	(1); (2); (4); (7)	(1), (2) e (7) reuniões presenciais; (4) consulta (reunião de apresentação e discussão do documento final).
Coleta de evidências	(1); (2); (3)	(1), (2) e (3) apoio logístico.
Interpretação e discussão de resultados	(1); (2); (3); (4); (5); (6)	Reuniões presenciais para discussão dos resultados da avaliação.
Planejamento para ação	(1); (2); (3); (4)	Reuniões presenciais para discutir o uso dos resultados da avaliação.

O propósito principal da avaliação foi o de apoiar o aprendizado das equipes do Canal Futura e da FRM e trazer contribuições que permitissem o aperfeiçoamento do trabalho realizado, para ampliar o impacto da intervenção. A relevância dos resultados do projeto para o público-alvo, as qualidades e limitações de sua implementação, a importância e a sustentabilidade dos resultados

produzidos e os diferentes aspectos (contextuais, metodológicos etc.) que potencializaram ou limitaram o nível de sucesso constam entre as questões avaliativas definidas pela equipe, listadas a seguir:

1. Em que medida o projeto contribui para atender necessidades e prioridades das organizações parceiras e dos potenciais beneficiários finais (educadores, crianças, famílias e comunidade)?
2. Quão bem o projeto foi desenhado e implementado?
 - 2.1. As estratégias e critérios adotados para selecionar as organizações parceiras foram os mais acertados para identificar as organizações mais efetivas para implementar o projeto?
 - 2.2. Os apoios iniciais e ao longo do projeto oferecidos pela equipe do Canal Futura às organizações parceiras (Maleta Completa e Maleta Básica) foram adequados para garantir o atingimento de resultados significativos?
 - 2.3. O apoio inicial e ao longo do projeto oferecido pelas organizações parceiras que receberam a Maleta Completa às demais organizações que receberam a Maleta Básica foi adequado para o atingimento de resultados significativos?
 - 2.4. Os mecanismos de monitoramento presencial e à distância adotados e o uso dos resultados desse monitoramento foram adequados para a boa execução do projeto?
3. Qual a qualidade e a importância dos resultados alcançados pelo projeto junto aos educadores, crianças, famílias/comunidade e organizações parceiras?
 - 3.1. Quão importante foi a contribuição do projeto para aumentar o conhecimento e melhorar o trabalho dos educadores nas escolas e em outros espaços educativos junto a

crianças de 0 a 11 anos e seus familiares ou a comunidade onde atuam?

- 3.2. O projeto contribuiu de forma significativa para o fortalecimento das organizações parceiras?
4. Onde (por exemplo: zona urbana *versus* rural; Nordeste *versus* Sudeste), para quem (por exemplo: ONGs *versus* governo; grandes *versus* médias/pequenas organizações parceiras; educadores infantis *versus* educadores do Ensino Fundamental) e por que o Projeto Maleta conseguiu ou não produzir resultados relevantes?
5. Os resultados justificaram os custos incorridos na elaboração e na distribuição da Maleta Infância e o tempo e outros recursos investidos pela equipe do Canal Futura e das organizações parceiras para sensibilização, apoio e monitoramento do projeto?
6. Qual a probabilidade de as mudanças positivas provocadas pelo projeto perdurarem após a finalização das ações planejadas pela equipe do Canal Futura?

Para que cada pergunta avaliativa pudesse ser efetivamente respondida, foram definidos os critérios de avaliação e as evidências que deveriam ser coletadas sobre cada um dos critérios ao longo da implantação. A fim de criar parâmetros que pudessem orientar a definição de quão boa foi a *performance* do projeto em cada uma das perguntas avaliativas, foram criadas rubricas avaliativas, que mostram como as evidências devem aparecer em cada nível de performance (excelente, bom, razoável e ruim). As rubricas foram desenvolvidas pelas equipes de Mobilização Comunitária do Canal Futura, do nível central e regional, pela equipe de Desenvolvimento Institucional da FRM e pelo consultor externo que facilitou o desenho da avaliação.

A avaliação utilizou abordagem mista, lançando mão de metodologia quantitativa e qualitativa, prevendo o acompanhamento de todo o ciclo de implementação do Projeto Maleta Infância. As

técnicas para coleta de dados incluíram: (i) entrevistas presenciais e telefônicas junto a diferentes grupos (educadores, organizações parceiras, mobilizadores comunitários, e equipe da FRM); (ii) pareceres de especialistas; (iii) revisão de documentos e outras evidências relevantes sobre o projeto e as organizações envolvidas; e (iv) estudos de caso de uma amostra intencional de organizações que tiveram maior e menor sucesso na implementação do projeto.

Entre os principais desafios decorrentes da opção feita pela nova metodologia estão: (i) o tempo necessário para a construção da avaliação, já que o processo pressupôs o envolvimento de vários atores em diferentes momentos; e (ii) a necessidade de estruturar uma estratégia de coleta de dados que fosse financeiramente viável e capaz de lidar com a complexidade e abrangência (inclusive geográfica) do projeto. Nesse sentido, privilegiou-se a realização de um número maior de entrevistas telefônicas em lugar de entrevistas presenciais.

Na visão da FRM, a metodologia adotada tem permitido a realização de um processo avaliativo orientado pelas questões que realmente importam para o grupo, o que viabiliza que sejam tiradas conclusões avaliativas a respeito do programa a partir de evidências concretas.

O uso de métodos mistos na avaliação do Programa Primeiríssima Infância

FUNDAÇÃO MARIA CECÍLIA SOUTO VIDIGAL

1. A FUNDAÇÃO MARIA CECÍLIA SOUTO VIDIGAL

Desde 2006, a Fundação Maria Cecília Souto Vidigal (FMCSV) atua na promoção do desenvolvimento da primeira infância por meio de um conjunto de ações integradas, voltadas para a geração e a disseminação do conhecimento prático e científico sobre o tema. Essa orientação se traduz em projetos que vão da intervenção social em municípios do estado de São Paulo a parcerias com centros de excelência no Brasil e no exterior; no incentivo a pesquisas; e na articulação de especialistas em desenvolvimento infantil com agentes que estão no centro da tomada de decisão da esfera pública, entre outras atividades.

Nosso trabalho se apoia no conceito de desenvolvimento integral e integrado da criança, ou seja, no entendimento de que ela deve ser vista como um todo, para que consiga atingir a plenitude do seu potencial. Isso significa que precisa das condições adequadas para se desenvolver física, emocional, social e culturalmente, e que essas dimensões interdependentes não podem ser tratadas de forma isolada. Todos são responsáveis por dar à criança os re-

cursos necessários para o seu desenvolvimento, dos pais aos gestores públicos, atuando conjuntamente, somando e compartilhando conhecimentos e estimulando seu crescimento de forma saudável e estruturada.

2. A AVALIAÇÃO NA FUNDAÇÃO MARIA CECÍLIA SOUTO VIDIGAL

A avaliação do impacto de suas iniciativas sempre fez parte de todas as ações da FMCSV, de tal modo que, em 2011, a organização decidiu criar uma área dedicada exclusivamente à avaliação e pesquisa. A área é responsável pela concepção dos processos de monitoramento e avaliação dos projetos e programas, pela realização de pesquisas aplicadas para gerar informações úteis à definição de estratégias da FMCSV, e para informar e mobilizar seus públicos de interesse.

O escopo das atividades de avaliação das práticas dos programas vai além dos aspectos tradicionais de gestão e aprendizagem com a implantação de projetos. As ações envolvem também a produção e gestão do conhecimento sobre programas e projetos voltados ao desenvolvimento na primeira infância. Internamente, a FMCSV acredita que, quanto mais uma organização avalia suas práticas e estratégias, mais ela aprende, se desenvolve e se torna efetiva. A prática da avaliação, uma vez inserida nas estratégias, políticas e procedimentos, passa a ser parte da cultura organizacional.

3. O PROGRAMA PRIMEIRÍSSIMA INFÂNCIA

Ao propor e implementar o Programa Primeiríssima Infância em diversos municípios do estado de São Paulo, a Fundação Maria Cecília Souto Vidigal (FMCSV) teve o objetivo de conceber um conjunto de estratégias para aprimorar as práticas de atendimento e cuidado à criança pequena e à gestante, por meio da disseminação de conhecimento sobre o Desenvolvimento da Primeira Infância (DPI) junto aos profissionais de unidades de saúde, Educação Infantil e assistência social, e pelo estímulo à mobilização comunitária e apoio à gestão e liderança.

Esses conhecimentos propiciam uma visão integral e integrada da criança como sujeito de direitos, especialmente quanto ao desenvolvimento pleno e saudável. O programa representa grandes desafios tanto para a FMCSV como para os municípios que aceitaram trilhar esse caminho.

O programa foi planejado em 2008 a partir de uma série de estudos na área de desenvolvimento infantil realizados por pesquisadores e consultores que compunham um comitê criado na FMCSV. Já naquela etapa, a avaliação foi identificada como um componente essencial no apoio ao planejamento da iniciativa. Dois avaliadores foram convidados para apoiar a definição de indicadores e critérios orientadores da seleção dos municípios que seriam contatados, o que foi feito a partir de dados secundários (Datusus, IBGE e Edudata) relacionados a aspectos-chave do desenvolvimento infantil. A partir dos critérios, foram selecionados oito municípios convidados a apresentar propostas para o programa. Ao final, seis municípios foram escolhidos e as ações foram iniciadas. O diálogo e a articulação com as novas administrações municipais foram as primeiras atividades.

O programa parte da premissa de que a sociedade tem conhecimentos tácitos sobre desenvolvimento infantil, especialmente na etapa do 0 aos 3 anos. Estudos apontam que tal conhecimento não é suficiente para a promoção do desenvolvimento infantil, por estar centrado em aspectos biológicos como o crescimento, o peso etc. Além desses aspectos, a ciência tem demonstrado que as dimensões emocionais, cognitivas e sociais são fundamentais para o desenvolvimento integral das crianças. Existem evidências de que o vínculo entre a gestante/mãe e o feto/bebê é muito importante para o bom desenvolvimento do cérebro. Poucos sabem que a aceitação da gravidez pela mãe e pelas pessoas que interagem com ela é aspecto crítico para a criação e o fortalecimento desse vínculo, possibilitando o melhor desenvolvimento de certas estruturas cerebrais do feto que, adiante, afetarão funções operativas.

Nesse contexto localiza-se o desafio central do programa: melhorar as ações de promoção do desenvolvimento na primeiríssima infância, período que se inicia na gravidez e perdura até o terceiro ano de vida das crianças, nos âmbitos político, dos serviços de aten-

dimento, familiar e comunitário, com base nas novas evidências da ciência nesse campo. Os objetivos principais do programa são:

- qualificar o atendimento das gestantes e crianças de 0 a 3 anos;
- mobilizar as comunidades para a promoção da primeira infância;
- desenvolver a governança local para construir políticas públicas integradas;
- levantar boas práticas de promoção do desenvolvimento infantil;
- disseminar o conhecimento gerado.

Para alcançar esses objetivos, lançou-se mão de três estratégias centrais:

- capacitações e formações de reeditores locais de conhecimentos e práticas;
- seminários de aprendizagem entre os municípios;
- perspectiva de avaliação como ferramenta de devolutiva (*feedback*) ao longo do programa.

Com eixos de intervenção voltados a profissionais, serviços, família, comunidade e políticas públicas, o primeiro ciclo do programa teve duração de quatro anos (2009-2012), visando cobrir todos os anos de uma gestão municipal. Um novo ciclo foi iniciado em 2011, dentro de uma perspectiva de ganho de escala e com novo formato, inclusive na avaliação. *Inputs* diretos nos municípios perderam intensidade, assim como a expectativa de resultados. Foi desenvolvida uma estrutura de governança regional do programa.

Desde então, parcerias com a Universidade de Harvard e com outras instituições, que se dedicam a estudar o *scaling up*, alimentam mudanças no programa, visando alcançar mais municípios, sem perder a qualidade nos aspectos mais relevantes. Nesse novo ciclo, os principais grupos de atores/parceiros envolvidos no programa são: (i) profissionais (técnicos) de saúde, educação e assistência – beneficiários de ações de capacitação e formação de multiplicadores; (ii) comércio e mídia local – meios de divulgação

de ações do programa abertas na cidade; (iii) sociedade em geral – beneficiária de eventos abertos, tais como campanhas e ações educativas; (iv) gestores de secretarias e outros atores do executivo (prefeituras) – responsáveis pelo alinhamento de estratégias locais e parcerias para realização das ações do programa junto às equipes necessárias e à sociedade em geral.

4. A POTÊNCIA, OS LIMITES E OS DESAFIOS EM AVALIAR O PROGRAMA

Após a etapa de seleção de projetos, foram planejados a sistemática e os instrumentos de marco zero, monitoramento e avaliação de resultados e impacto. Nesse processo foi elaborada a teoria de mudança (*Theory of Change*) do programa, que definia suas estratégias centrais e o que deveria ser transformado no cenário do desenvolvimento infantil nos municípios e os impactos que se poderia esperar.

A abordagem adotada pela avaliação dispensou a construção de perguntas avaliativas, mas trouxe preocupações em conhecer em que medida a teoria de mudança fora efetivada, bem como os avanços dos municípios em resultados desejados, além dos previstos. A avaliação carregava, assim, dois sentidos principais: buscar evidências em relação à teoria de mudança e servir como subsídio para a gestão ao longo do ciclo de vida dos projetos locais e do programa como um todo.

O processo de amadurecimento dos indicadores deu-se por todo o ciclo do projeto. Buscou-se identificar e selecionar apenas aqueles que fossem sensíveis à realidade e que traduzissem de fato as estratégias essenciais do programa. De um conjunto inicial de 139 indicadores, chegou-se a 50, após três anos de implantação. A distribuição dos indicadores em oito dimensões manteve-se em todo o processo, a saber:

- profissionais preparados;
- pré-natal adequado;
- puericultura adequada;
- parto e puerpério adequados;

- Educação Infantil adequada;
- espaços lúdicos adequados;
- mães e pais preparados; e
- comunidade mobilizada.

Exemplos de indicadores dentro das dimensões:

- a. pai ou companheiro(a) acompanha o pré-natal de sua mulher;
- b. profissionais estimulam e possibilitam a participação do pai/outro(a) acompanhante no parto;
- c. encontros de grupos de “famílias grávidas” oferecem apoio específico para adolescentes grávidas.

O desenho metodológico buscou contemplar a possibilidade de estar sensível ao que o campo pudesse trazer e revelar, sem se fixar rigidamente nos indicadores assumidos no início. Além de avaliar os ganhos de capacidade dos profissionais, a melhoria dos serviços de atenção a crianças, as mudanças nas políticas públicas e a ampliação da mobilização comunitária em torno da primeira infância, considerou-se importante avaliar o desenvolvimento das crianças e também como identificar e compreender mais amplamente o impacto das mudanças.

Após ampla pesquisa sobre alternativas para avaliar o desenvolvimento das crianças na primeira infância, a equipe da FMCSV identificou o *Early Development Instrument* (EDI), um instrumento desenvolvido pelo *Offord Centre for Child Studies*, ligado à Universidade McMaster, em Ontário, Canadá. Ele é indicado para avaliar aspectos do desenvolvimento infantil no final da primeira infância, podendo ser aplicado em crianças entre 4 e 6 anos. Propõe-se a medir a maturidade e o grau de prontidão delas em cinco domínios críticos para o ingresso na vida escolar.

O EDI vem sendo aplicado no Canadá e na Austrália como um instrumento para avaliação em escala nacional. Seus resultados são utilizados para orientar políticas públicas para a primeira infância. O levantamento das informações é feito por meio do preenchimento de questionários por educadores, tomando como base sua observação dos comportamentos de seus alunos em cinco domínios:

1. bem-estar físico;
2. competência social;
3. saúde e maturidade emocional;
4. desenvolvimento cognitivo e linguagem;
5. capacidade de comunicação e conhecimentos gerais.

O EDI contempla 120 questões que buscam identificar em que medida a família, a escola e a comunidade são capazes de estimular o desenvolvimento de suas crianças até o ingresso na escola, além de indicar o grau de preparo delas para conviver no ambiente escolar e usufruir dele. O EDI não é usado como avaliação individual de uma criança, mas como diagnóstico populacional aplicado em todas as crianças de determinada comunidade.

Em parceria com o Programa Infância Melhor (PIM)¹ e o *Offord Center*, a FMCSV promoveu a tradução de todo o material do EDI para o português e sua adaptação à realidade brasileira. O resultado do esforço foi um EDI validado para o contexto brasileiro, a partir de um estudo piloto nos seis municípios parceiros da FMCSV. O instrumento foi aplicado em 2011 para o estabelecimento de uma linha de base em todas as crianças de 5 anos que frequentam os centros de Educação Infantil nesses municípios e em algumas unidades de Educação Infantil da Microrregião de Cidade Ademar, na cidade de São Paulo.

Esse estudo servirá de base de comparação com um próximo estudo, a ser feito em 2015, que avaliará o impacto no grau de maturidade e de prontidão das crianças que foram beneficiadas pelo Projeto de Intervenção Local da FMCSV. Além de servir como parâmetro para a avaliação de impacto dos projetos sociais, seus resultados poderão ser utilizados para orientar políticas locais para a primeira infância. >

1. O PIM integra a política de governo do estado do Rio Grande do Sul, sob a coordenação da Secretaria da Saúde e apoio das secretarias da Educação, Cultura, Trabalho e Desenvolvimento Social. Desenvolvido desde 2003, tornou-se Lei Estadual nº 12.544 em 03 de julho de 2006. É um programa institucional de ação socioeducativa voltado às famílias com crianças de zero até seis anos e gestantes, que se encontram em situação de vulnerabilidade social.

ETAPAS DA AVALIAÇÃO NO CICLO DE VIDA DO PROGRAMA PRIMEIRÍSSIMA INFÂNCIA																				
Atividades de Avaliação	2009		2010				2011				2012				2013				2014	2015
	3º	4º	1º	2º	3º	4º	1º	2º	3º	4º	1º	2º	3º	4º	1º	2º	3º	4º		
Linha de base																				
Devolutivas de linha de base																				
Monitoramento																				
Devolutivas do monitoramento																				
Aplicação do inventário <i>Early Development Instrument</i> (EDI)																				
Avaliação de resultados parciais																				
Devolutivas de resultados parciais																				
Avaliação de impacto																				
Devolutivas de avaliação de impacto																				
Reaplicação do EDI																				

Fonte: Fundação Maria Cecília Souto Vidigal.

A avaliação envolveu diversos atores. Na coleta de informações, além dos próprios avaliadores, foram envolvidos pesquisadores contratados e educadores de algumas unidades escolares na aplicação do EDI. Um grande público formado por profissionais de diversas áreas (técnicos e gestores de saúde, Educação e assistência, principalmente) participou das devolutivas, dentro de um processo aberto à participação e à construção de conhecimentos e de aprendizagens sobre o programa de maneira colaborativa.

O modelo avaliativo misto utilizado para o programa, associando o instrumento quantitativo (EDI), aplicado censitariamente às crianças, grupos focais e entrevistas individuais, bem como questionários e fontes secundárias, teve elevada capacidade de

aprofundar a busca de evidências de resultados em relação aos impactos sugeridos pelos indicadores. O modelo favoreceu a produção, a partilha e o uso de achados da avaliação pelos municípios parceiros, e também a análise e redesenho de estratégias de expansão do programa pela FMCSV.

Outros potenciais podem ser ainda apontados nos seguintes aspectos:

- a. o programa construiu um claro modelo da mudança que se queria implementar (teoria de mudança);
- b. o programa realizou uma avaliação de marco zero;
- c. o programa construiu indicadores recomendados e reconhecidos por especialistas, fortalecendo-se no campo do desenvolvimento infantil e facilitando a apresentação e o debate sobre o tema;
- d. a avaliação implementou devolutivas cuidadosas, buscando verificar o quanto os resultados eram condizentes com o que realmente estava acontecendo, ou se sub ou superestimava a realidade, favorecendo o amadurecimento e a calibragem dos indicadores; e
- e. a avaliação lançou mão de formas gráficas para apresentar os resultados, facilitando o diálogo com os públicos externo e interno do programa.

Quanto aos principais desafios enfrentados pela FMCSV, podem ser resumidos da seguinte maneira:

- a. no fato da avaliação ter consumido cerca de 20% do orçamento total do programa;
- b. na dificuldade de transmitir ao grupo de técnicos que estruturou o programa, a importância e a capacidade de um modelo de avaliação de natureza não experimental, com orientação qualitativa, mas baseado em evidências capazes de apontar para o entendimento dos resultados;
- c. no esforço para identificar indicadores de dados secundários sensíveis que pudessem dialogar com o programa e os resultados previstos nas cidades;

- d. no trabalho com o Conselho da FMCSV, com a perspectiva de não se ter metas quantitativas rígidas estabelecidas nas cidades em todas as frentes, dado que em alguns casos isso não era possível; e
- e. na negociação intensa com prefeitura e educadores em relação ao tempo necessário para a aplicação do EDI, por conta do impacto na rotina de trabalho gerado (de 12 a 15 horas demandadas de cada educador) para o preenchimento dos instrumentos de todas as crianças de uma turma de crianças (20 a 25).

Quanto aos limites do modelo, eles podem ser percebidos, primeiro, pela composição dos 50 indicadores que concentram um maior número de indicadores qualitativos que não permitem o estudo da extensão dos resultados nas unidades de Educação ou de saúde específicas nas cidades. Em complemento, outro limite esteve na forma como os resultados do programa nas dimensões e nos municípios foram calculados e apresentados, agregando números e usando médias que nem sempre representam um parâmetro preciso dos alcances.

A utilidade complementar da avaliação, além dos insumos gerados pelo próprio processo na gestão do programa, foi a geração de conteúdos para a gestão de conhecimento na FMCSV. A fundação fez uso de dados e de resultados para influenciar a formação de opinião, divulgando *briefings* para gerar matérias jornalísticas sobre o tema (desenvolvimento infantil) em mídias diversas, inclusive em cada um dos municípios parceiros. Foi gerado um documento de registro dos principais resultados e bases conceituais, que tem favorecido diálogos com outras prefeituras e organizações que trabalham para o desenvolvimento infantil.

Abordagens quantitativas na avaliação de impacto do Programa Jovens Urbanos

FUNDAÇÃO ITAÚ SOCIAL

1. A FUNDAÇÃO ITAÚ SOCIAL

A Fundação Itaú Social, uma das organizações responsáveis por realizar o investimento social do Itaú, tem como atividades centrais a formulação, a implantação e a disseminação de metodologias voltadas à melhoria de políticas públicas na área educacional e à avaliação econômica de projetos sociais.

A atuação da Fundação abrange todo o território brasileiro, desenvolve-se em parceria com governos, setor privado e organizações da sociedade civil e é orientada pelo foco em resultados, buscando contribuir de maneira efetiva para a transformação social.

2. A AVALIAÇÃO NA FUNDAÇÃO ITAÚ SOCIAL

A constante preocupação em relação aos resultados atingidos pelas escolhas inerentes aos programas em que atua, assim como o questionamento sempre presente a respeito dos impactos sociais efetivamente gerados pelo investimento realizado, revela a compreensão da avaliação como um instrumento fundamental para qualificar e aprimorar o investimento social realizado pela Fundação Itaú Social.

Uma gestão de projetos fundamentada na efetividade dos programas exige princípios de atuação e práticas avaliativas que tragam respostas objetivas sobre os resultados alcançados frente aos recursos investidos, além de contribuir com a transparência, a gestão e a prestação de contas junto a todas as partes interessadas. Refletida em sua missão, as práticas de avaliação são uma importante ferramenta na busca das respostas necessárias para a condução e o aprimoramento das ações sociais realizadas pela Fundação Itaú Social.

3. O PROGRAMA JOVENS URBANOS

Na sua concepção original, o Programa Jovens Urbanos teve como principal objetivo ampliar o repertório sociocultural na perspectiva da Educação Integral de jovens entre 16 e 21 anos em situação de vulnerabilidade social em contextos urbanos. O programa tinha duração de 16 meses e realizava atividades no contraturno escolar, com foco em processos que possibilitassem o desenvolvimento de competências e habilidades dos jovens para a vida pública e pessoal, colaborando para a qualificação do seu acesso ao mundo do trabalho. Além disso, estimulava o acesso aos equipamentos e bens culturais e sociais disponíveis na cidade e incentivava a permanência dos jovens na escola e sua vinculação a novos processos formativos.

Desde a sua primeira edição, o programa atua em rede: é iniciativa da Fundação Itaú Social com coordenação técnica do Cenpec¹, execução por ONGs locais e ações de formação desenvolvidas em parceria com o poder público, fundações, institutos e assessores tecnológicos, corresponsáveis pela produção de condições de aprendizagem qualificadas e pelo alcance dos resultados pretendidos.

O programa assume diferentes formatos conforme o município e parceiro que o implementa, e sua metodologia está centrada em quatro eixos vinculados ao direito à cidade e ao desenvolvi-

1. Centro de Estudos e Pesquisas em Educação, Cultura e Ação Comunitária.

mento integral: (1) juventudes e culturas urbanas; (2) juventudes e tecnologias contemporâneas; (3) juventudes e mundo do trabalho; e (4) juventude e participação política. A estratégia de formação deriva de quatro concepções de aprendizagem: exploração, experimentação, produção e expressão.

O objetivo dessa metodologia é promover o desenvolvimento de competências e o envolvimento reflexivo dos jovens com os territórios da cidade, oferecendo oportunidades para que conheçam e explorem espaços onde estão concentradas práticas juvenis, artísticas, tecnológicas, relativas ao mundo do trabalho, das políticas, das ciências, de promoção da saúde, de lazer e de esportes.

4. A AVALIAÇÃO DO PROGRAMA

O propósito da avaliação do Programa Jovens Urbanos foi conhecer o impacto dessa iniciativa na vida dos jovens participantes nos aspectos (1) inserção no mercado de trabalho; (2) frequência escolar; (3) renda pessoal; (4) envolvimento com a política; (5) leitura de livros e revistas; e (6) escolaridade.

Houve quatro processos avaliativos, acompanhando a primeira e a terceira edições do programa. A referência metodológica para esses estudos foi a econométrica, fundamentada em princípios de estatística e econometria, em consonância com a cultura organizacional da Fundação Itaú Social, que realiza avaliação econômica de seus investimentos sociais, sendo esta composta pela avaliação de impacto e retorno econômico.

A avaliação de impacto permite quantificar relações causais entre as ações do projeto e os indicadores de interesse, permitindo afirmar se foi de fato o projeto o elemento responsável pelas alterações ocorridas, ou seja, se foram realmente resultado da intervenção ou se ocorreriam de qualquer forma. O retorno econômico mede o retorno do programa em termos monetários para a sociedade, por meio da comparação dos custos do projeto com os benefícios gerados para o público atendido.

Essa abordagem de avaliação exige a definição de um grupo de tratamento, sendo este constituído dos indivíduos que recebem uma intervenção, e outro de controle, que reúne indivíduos com

características semelhantes àqueles do grupo de tratamento, sem, contudo, receberem a intervenção. O desempenho dos grupos é comparado de tal forma que as diferenças em seus resultados podem ser associadas ao programa.

A primeira edição do Jovens Urbanos ocorreu nos bairros de Campo Limpo, zona sul de São Paulo, e Brasilândia, zona norte de São Paulo, regiões com altos índices de vulnerabilidade juvenil, o que pode ser comprovado pela leitura de indicadores produzidos pela Fundação Seade². Ao se inscreverem no processo seletivo, os candidatos realizaram uma prova de habilitação em matemática, para identificação do nível de conhecimento mínimo que tinham nessa disciplina.

Dentre esses jovens, 960 foram considerados habilitados para o programa e formaram o universo para compor o sorteio que iria, de maneira aleatória, definir os grupos de tratamento e de controle. Parte do grupo de jovens habilitados participaria do programa, enquanto outra parte não o faria, sendo contudo acompanhada pelo processo de avaliação. Todos os 960 jovens foram incluídos no cadastro do Programa Pró-Social da Secretaria Estadual de Assistência e Desenvolvimento Social, para que recebessem a bolsa Ação Jovem da Secretaria. Contudo, o grupo tratamento recebeu, além da bolsa, a intervenção do Programa Jovens Urbanos.

Nesse contexto, a primeira avaliação, realizada em 2006, teve amostra de 480 jovens no grupo de tratamento e a mesma quantidade no grupo de controle. A amostra da segunda avaliação, realizada em 2010, e que buscou os mesmos jovens, foi de 329 jovens para o grupo de tratamento (53%) e 294 para o grupo de controle (47%).

Um ano após o término das intervenções do programa, os jovens dos dois grupos foram convidados a responder um questionário no qual se buscava conhecer a existência de melhora de vida nos seguintes aspectos relacionados à formação proporcionada pelo programa:

2. Fundação Sistema de Análise de Dados, ligada à Secretaria de Planejamento e Desenvolvimento Regional do Governo do Estado de São Paulo.

- inserção no mercado de trabalho;
- frequência escolar;
- renda pessoal;
- envolvimento com a polícia;
- leitura de livros e revistas.

Para estimar eventuais diferenças de resultados entre os dois grupos, foram comparados seus desempenhos em relação a essas variáveis. No caso do sorteio, a metodologia de avaliação nos garante que as diferenças entre os resultados dos grupos pode ser lida como impacto. Os resultados indicaram que houve diferença entre os grupos nos aspectos relacionados ao mercado de trabalho, renda pessoal e envolvimento com a polícia; todos favoráveis aos participantes e de acordo com as expectativas iniciais para os resultados do programa.

A segunda avaliação, realizada três anos após o término das ações, entretanto, indicou que o programa não gerou efeitos duradouros sobre a vida daqueles jovens, o que significa que o impacto foi de curto prazo, diluindo a diferença observada na primeira avaliação, entre participantes e não participantes, ao longo do tempo. Os resultados das avaliações orientaram os ajustes abaixo na formatação das novas edições do Programa Jovens Urbanos: >

Dimensão	Ajustes
Renda e empregabilidade	<ul style="list-style-type: none"> • Inclusão do eixo mundo do trabalho • Qualificação do monitoramento da evasão do programa gerada por inserção no mercado de trabalho
Hábitos de leitura	<ul style="list-style-type: none"> • Reforço no eixo letramento • Continuidade das ações de expressão • Criação da semana Expressão Jovens Urbanos
Escolaridade formal	<ul style="list-style-type: none"> • Reforço no monitoramento de jovens que se evadem da escola • Programa ocorrendo dentro das escolas em parceria com Secretarias de Educação
Ensino Técnico Profissionalizante	<ul style="list-style-type: none"> • Parcerias com escolas técnicas
Ensino Superior	<ul style="list-style-type: none"> • Parceria com universidades

A terceira edição do Programa Jovens Urbanos, iniciada em 2007 e finalizada em 2008, passou por alguns ajustes fundamentados em informações de monitoramento e avaliação de resultado

das duas primeiras edições, somadas às aprendizagens identificadas pela avaliação econômica da primeira edição.

Nela, os grupos de participantes da avaliação foram definidos a partir das inscrições realizadas no programa. Nesse caso, o grupo de controle foi selecionado a partir da lista de espera do programa, sem uma seleção aleatória dos participantes, caracterizando o desenho como quase-experimental. A amostra foi composta por 185 jovens no grupo de controle e 366 no grupo de tratamento. Na ausência do sorteio, foi utilizado um método de pareamento para garantir que a comparação de resultados fosse feita entre grupos semelhantes. A ideia era atribuir pesos diferenciados aos jovens do grupo de controle de acordo com características e probabilidades de participação no programa para criar um grupo mais parecido com o dos beneficiários.

Na segunda avaliação da terceira edição a amostra foi de 122 jovens do grupo de controle e 254 jovens do grupo de tratamento. Os resultados indicaram que houve impacto de curto prazo na vida dos jovens nos aspectos renda pessoal, mercado de trabalho, leitura de jornal e participação em ONGs. Em 2012, dois anos após a primeira avaliação, a avaliação com os mesmos grupos indicou impacto positivo de longo prazo nos aspectos ingresso no Ensino Superior e na escolaridade. O quadro a seguir sintetiza o desenho adotado pela avaliação econômica na primeira e na terceira edições do programa: >

PROCESSO AVALIATIVO DO PROGRAMA JOVENS URBANOS				
	Primeira edição		Terceira edição	
	Primeira avaliação	Segunda avaliação	Primeira avaliação	Segunda avaliação
Data	2006	2010	2010	2012
Grupo de tratamento	Aleatório Amostra: 480	Aleatório Amostra: 329	Não aleatório Amostra: 366	Não aleatório Amostra: 254
Grupo de controle	Aleatório Amostra: 480	Aleatório Amostra: 294	Não aleatório Amostra: 185	Não aleatório Amostra: 122

Os principais desafios que a avaliação enfrentou na utilização desse método foram: (a) a dificuldade de encontrar os jovens que participaram do programa nos anos anteriores para verificar o im-

pacto de longo prazo em suas vidas; (b) a desistência de participação dos jovens sorteados ainda na primeira edição do programa, que trouxe limitações ao uso do método aleatório, exigindo a utilização de métodos de estimação mais sofisticados que a simples comparação de médias.

Importante também ressaltar que metodologias complementares de avaliação estão presentes em todas as etapas do programa. Processos estruturados de monitoramento do planejamento e das ações desenvolvidas, associados a diferentes avaliações de resultados ao final das atividades realizadas e somados às avaliações de impacto, compõem um ferramental analítico extremamente valioso e útil para o constante aperfeiçoamento do Programa Jovens Urbanos.

Após avaliações sistemáticas de diferentes naturezas, o programa foi aperfeiçoado em relação a critérios de idade do público participante, atualmente de 15 a 20 anos; à duração da edição, atualmente de 12 meses; e a intervenções do programa em escolas públicas do território.

Hoje o programa está na oitava edição em São Paulo (SP) e na segunda edição em Serra (ES) e em Pouso Alegre (MG). Desde sua primeira edição, alcançou aproximadamente 7 mil jovens. As modificações até aqui produzidas no programa, com base nas avaliações econométricas, qualitativas e no monitoramento, têm sido essenciais para aperfeiçoar a iniciativa e ampliar seu potencial de produzir resultados positivos na vida dos jovens participantes.

Oportunidades para atualização e publicação na área de avaliação

A Fundação Roberto Marinho e o Instituto Fonte realizaram um levantamento sobre oportunidades de formação, publicação (em revistas direcionadas à avaliação e outras acadêmicas) e articulação para aqueles que estiverem interessados em aprofundar seu contato com a avaliação, tanto no Brasil quanto na América Latina e em outros continentes.

O objetivo é contribuir para o fortalecimento da troca e da formação entre pessoas interessadas no campo. Nos quadros a seguir, estão identificadas tais iniciativas, seguidas de seus endereços eletrônicos na Internet, e de pequenas explicações sobre sua natureza, periodicidade etc. Esses textos explicativos foram construídos com base nos sites pesquisados e por isso reproduzem suas informações. Nesse sentido, não fazemos referências bibliográficas específicas em cada caso e afirmamos que as citações entre aspas são transposições diretas dos conteúdos dos sites.

Recomendamos que os leitores visitem as iniciativas de seu interesse para maiores esclarecimentos. >

OPORTUNIDADES DE FORMAÇÃO					
Nome	Tipo	Disponível em *	Organização	Descrição	
Curso de Especialização de Avaliação em Saúde	Especialização a distância	http://www.ead.fiocruz.br/cursos/index.cfm?cursoId=1574	Escola Nacional de Saúde Pública (ENSP/ Fiocruz)	O curso é parte da estratégia nacional de institucionalização do Monitoramento e Avaliação (M&A) no Sistema Único de Saúde. Está voltado para o desenvolvimento de habilidades e competências necessárias à prática do avaliador, no sentido da promoção da melhoria de qualidade das práticas em saúde e da transformação social. O curso faz parte de uma ação interinstitucional, de caráter inovador, promovida pela Escola Nacional de Saúde Pública Sérgio Arouca, da Fundação Oswaldo Cruz, por meio do Laboratório de Avaliação de Situações Endêmicas Regionais, do Departamento de Endemias Samuel Pessoa, e da Coordenação de Educação a Distância. A ação desenvolve-se em parceria com a Central de Monitoramento e Avaliação do Departamento de DST/Aids e Hepatites Virais da Secretaria de Vigilância em Saúde do Ministério da Saúde e conta com o apoio do Center for Disease Control and Prevention.	
Programa de Pós-Graduação em Avaliação de Políticas Públicas da Universidade Federal do Ceará (UFC)	Mestrado profissional	http://www.mapp.ufc.br/	UFC	O Mestrado em Avaliação de Políticas Públicas da UFC, modalidade profissional, foi aprovado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) em 21/12/2004. Foi o primeiro curso de mestrado com área de concentração específica em avaliação de políticas públicas no País. Além do caráter inédito e inovador da proposta, vem constituindo produção científica na área desde 2005, em decorrência da defesa de dissertações – fato bem-vindo, dada a escassa literatura nacional especializada.	
Mestrado Profissional em Avaliação (stricto sensu)	Mestrado profissional	http://mestrado.cesgranrio.org.br/mestrado/objetivo.aspx	Fundação Cesgranrio	Tem como principal objetivo formar profissionais de avaliação capazes de planejar, conduzir, relatar e utilizar a avaliação de sistemas, programas, instituições e materiais quanto aos avanços teóricos, metodológicos e práticos da área.	
Coordenadoria de Estudos e Desenvolvimento de Projetos Especiais/Pontifícia Universidade Católica de São Paulo (Cedepe/ PUC-SP)	Grupo acadêmico	http://www.pucsp.br/cedepe/cedepe.html	PUC-SP	Grupo que tem como uma das áreas de atuação a avaliação. Segundo informações do site, "A construção de sistemas e matrizes de monitoramento e avaliação é uma aposta da Cedepe na perspectiva de contribuir para o fortalecimento da gestão pública. Nessa perspectiva, a Cedepe pode apoiar e reforçar o desenvolvimento institucional de organizações públicas e privadas para subsidiar a tomada de decisões sobre políticas, programas e projetos."	

* Último acesso em setembro de 2013.

OPORTUNIDADES DE FORMAÇÃO					
Nome	Tipo	Disponível em*	Organização	Descrição	
Master: Gestión, Analisis y Evaluación de Asuntos Públicos (e-learning)	Curso e-learning	http://www.iestqm.com/	TQM, Investigación y Formación, S.L./ Universidad de Alcalá de Henares, Espanha	Curso a distância voltado especialmente para gestão de políticas públicas oferecido pela Universidad de Alcalá de Henares, Espanha. Os objetivos são formar os participantes como analistas e avaliadores de políticas públicas; gerar uma rede de assessores que proponham melhorias no funcionamento das organizações; vincular a avaliação com a modernização administrativa e as Ferramentas da Nova Gestão Pública; ensinar técnicas e métodos para realizar uma análise e avaliação de qualidade; delimitar o papel da avaliação em sua contribuição para a melhoria da tomada de decisões e a reformulação de políticas públicas; conhecer boas práticas na avaliação de políticas públicas e contribuir com uma ética de desempenho comprometida com a equidade.	
International Program for Development Evaluation Training	Especialização presencial e a distância	http://www.ipdet.org/default.aspx	Apoiado pelo Independent Evaluation Group of the World Bank e pela Faculty of Public Affairs at Carleton University, Canadá	Centro de estudos e formação sediado em Ottawa, Canadá. Criado em 2001, o Programa Internacional de Treinamento para a Avaliação do Desenvolvimento (IPDET) é um programa de treinamento executivo que tem por objetivo dar a gestores e profissionais as ferramentas genéricas necessárias para avaliar políticas, programas e projetos de desenvolvimento em âmbito local, nacional, regional e global. O programa consiste em um curso de duas semanas, que cobre os aspectos essenciais, seguido por duas semanas de workshops aprofundados, ensinados por um grupo dinâmico de professores de renome internacional, de organizações dos hemisférios Norte e Sul. Uma iniciativa de treinamento intensiva e única do Grupo de Avaliação Independente do Banco Mundial e da Universidade de Carleton.	
Claremont Evaluation Centre (CEC)	Webinars	http://www.cgu.edu/pages/9962.asp	Institute of Organizational and Program Evaluation Research (IOPER)/ Claremont Evaluation Centre/Claremont University, Califórnia, EUA	Unidade de Pesquisa e Avaliação Organizadas dos Claremont Colleges. A missão do CEC (Centro de Avaliação Claremont, em português) é "prestar serviços e conduzir pesquisas e avaliações para melhorar a eficiência de uma ampla gama de projetos, programas, políticas e organizações". Usando conhecimento científico e metodologias de última geração, o CEC já prestou esse tipo de serviço para mais de 150 organizações diferentes nos últimos 15 anos. Vale a pena assistir aos webinars produzidos em 2012 (Claremont Evaluation Center Webinar Series: < http://www.cgu.edu/pages/10257.asp >).	
My M&E	Plataforma de partilha de informações e e-learning	http://mymande.org/	Unicef Evaluation Office and International Organization for Cooperation in Evaluation (IOCE)	Plataforma interativa da web para compartilhar conhecimentos nos sistemas nacionais de monitoramento e avaliação em todo o mundo. Além de ser uma fonte de aprendizado, promove o fortalecimento de uma comunidade global, enquanto identifica boas práticas e lições sobre monitoramento e avaliação em geral, e sobre os países que usam os sistemas M&E em particular.	

* Último acesso em setembro de 2013.

OPORTUNIDADES DE FORMAÇÃO				
Nome	Tipo	Disponível em *	Organização	Descrição
EvalPartners	Plataforma de partilha de informações e e-learning	http://www.mymande.org/evalpartners/	Unicef Evaluation Office e IOCE	Iniciativa vinculada ao My M&E com atenção especial às organizações da sociedade civil, ao fortalecimento dos VOPEs (Voluntary Organizations of Professional Evaluators, Organizações Voluntárias de Avaliadores Profissionais, em português) e à troca entre países. "O número de VOPEs está crescendo rapidamente mundo afora. Há grande espaço para a troca de soluções, ideias e experiências domésticas e voltadas para países em particular, para ajudar a capacidade de desenvolvimento de avaliação. Nesse contexto, a EvalPartners tem como objetivo fortalecer a capacidade de avaliação das organizações da sociedade civil por meio de cooperação triangular e do apoio dos profissionais uns aos outros e do aprendizado mútuo. A EvalPartners também aproveita as novas tecnologias de informação e comunicação para facilitar o networking e o aprendizado."
International Development Evaluation Association (Ideas)	Plataforma de partilha de informações e e-learning	http://conference.ideas-int.org/	Ideas	A Ideas (Associação Internacional de Avaliação do Desenvolvimento, em português) é uma associação independente de profissionais de avaliação, pessoas interessadas no tema e aqueles que buscam trazer informação e conhecimento com credibilidade para a agenda do desenvolvimento. A Ideas foi criada em setembro de 2002 em Pequim, na China. Está aberta a indivíduos e organizações de todo o mundo e, nesse momento, conta com quase 800 membros em mais de 100 países. A missão da Ideas é "melhorar e estender a prática de avaliação do desenvolvimento, aprimorando o conhecimento, fortalecendo capacidades e expandindo as redes para a avaliação do desenvolvimento, em particular nos países em desenvolvimento e em transição".
Monitoring and Evaluation News	Website	http://mande.co.uk/	Diversas	Website elaborado por Rick Davies, desde 1997, com suporte de diversas organizações inglesas (Oxfam, Action Aid, SCF and Christian Aid, entre outras). "Este site enfoca métodos de monitoramento e avaliação de progresso e resultado de programas de ajuda para o desenvolvimento. Muitos dos métodos discutidos aqui também são relevantes para a prática de avaliação quando aplicados a políticas e programas implementados por governos e organizações da sociedade civil em países ricos, não apenas em países que recebem ajuda internacional."
Plataforma Regional de Desarrollo de Capacidades en Evaluación y Sistematización de América Latina y el Caribe (Preval)	Plataforma de partilha de informações e e-learning	http://preval.org/es	Preval	"Preval é uma plataforma internacional – única em seu gênero – que assessora governos, unidades técnicas e organizações rurais para fortalecer sua capacidade de elaborar e desenvolver sistemas de Planejamento, Seguimento e Avaliação (PSA), como espaços de produção e informação para a tomada de decisões e a aprendizagem. Desde 1997, tem prestado serviço a mais de 30 organizações e 15 países da região. Tem como foco a questão do desenvolvimento rural. Além disso, publica boletim e oferece cursos a distância."

* Último acesso em setembro de 2013.

OPORTUNIDADES DE PUBLICAÇÃO – REVISTAS DIRECIONADAS À AVALIAÇÃO					
Nome	Tipo	Disponível em *	Organização	Descrição	
Estudos em Avaliação Educacional	Revista quadrimestral	http://www.fcc.org.br/biblioteca/apresenta_eae.php?area=publicacoes	Fundação Carlos Chagas	“Estudos em Avaliação Educacional publica trabalhos inéditos referentes à avaliação educacional, apresentados sob a forma de relatos de pesquisa, ensaios teóricos, metodologias, revisões críticas, artigos e resenhas.”	
Ensaio – Avaliação e Políticas Públicas em Educação	Revista trimestral	http://www.cesgranrio.org.br/publicacoes/principal.aspx	Fundação Cesgranrio	Trata-se de um veículo de divulgação de pesquisas, levantamentos, estudos, discussões e outros trabalhos críticos no campo da Educação, concentrando-se nas questões de avaliação e políticas públicas e enfatizando as experiências e perspectivas brasileiras.	
Revista Brasileira de Monitoramento e Avaliação	Revista semestral	http://ap.l.mds.gov.br/ojs/index.php/RBMA/index	Rede Brasileira de Monitoramento e Avaliação, Secretaria de Informação do Ministério do Desenvolvimento Social e Combate à Fome (SAGI/MDS) e da Secretaria de Assuntos Estratégicos da Presidência República (SAE/PR). Sua criação foi motivada pela necessidade de oferecer novos espaços para publicação de trabalhos no campo, cada vez mais frequentes em função do avanço da cultura de formulação, monitoramento e avaliação de políticas e programas. Além disso, a dispersão desse crescente volume de estudos dentre os diversos e consagrados periódicos já existentes nas ciências sociais brasileiras – cada um com seu viés disciplinar e sua comunidade acadêmica almejada – tem criado dificuldades para que os interessados possam ter uma visão mais integrada, multidisciplinar e holística da área de M&A. A revista procura, pois, criar oportunidades para o registro e intercâmbio de estudos de distintas naturezas – conceitual, metodológica e aplicada – e diferentes abordagens de pesquisa empírica – qualitativa, quantitativa, quase-experimental, estudos de caso etc.	A <i>Revista Brasileira de Monitoramento e Avaliação</i> é uma nova proposta editorial para a sistematização e disseminação de experiências e conhecimentos em Monitoramento e Avaliação (M&A) em políticas públicas, programas e projetos sociais no Brasil. É uma iniciativa de pesquisadores, técnicos e participantes da Rede Brasileira de Monitoramento e Avaliação (RBMA), com apoio da Secretaria de Avaliação e Gestão da Informação do Ministério do Desenvolvimento Social e Combate à Fome (SAGI/MDS) e da Secretaria de Assuntos Estratégicos da Presidência República (SAE/PR). Sua criação foi motivada pela necessidade de oferecer novos espaços para publicação de trabalhos no campo, cada vez mais frequentes em função do avanço da cultura de formulação, monitoramento e avaliação de políticas e programas. Além disso, a dispersão desse crescente volume de estudos dentre os diversos e consagrados periódicos já existentes nas ciências sociais brasileiras – cada um com seu viés disciplinar e sua comunidade acadêmica almejada – tem criado dificuldades para que os interessados possam ter uma visão mais integrada, multidisciplinar e holística da área de M&A. A revista procura, pois, criar oportunidades para o registro e intercâmbio de estudos de distintas naturezas – conceitual, metodológica e aplicada – e diferentes abordagens de pesquisa empírica – qualitativa, quantitativa, quase-experimental, estudos de caso etc.	
Revista Avaliação de Políticas Públicas	Revista semestral	http://www.mapp.ufc.br/wp-aval/institucional/sobre-a-aval	Sob responsabilidade do Mestrado Profissional em Avaliação de Políticas Públicas, da Universidade Federal do Ceará (UFC)	“A <i>Revista Avaliação de Políticas Públicas (Aval)</i> volta-se primordialmente a: publicação de análises e resultados de pesquisas em avaliação de políticas públicas; reflexões teórico-metodológicas sobre avaliação; desenvolvimento de ferramentas e estratégias metodológicas que contribuam para a avaliação de políticas públicas e reflexões sobre o exercício da multi e da interdisciplinaridade.”	

* Último acesso em setembro de 2013.

OPORTUNIDADES DE PUBLICAÇÃO – REVISTAS DIRECIONADAS À AVALIAÇÃO					
Nome	Tipo	Disponível em *	Organização	Descrição	
Avaliação: Revista da Avaliação da Educação Superior	Revista eletrônica	http://periodicos.uniso.br/ojs/index.php?journal=avaliacao&page=index	Rede de Avaliação Institucional da Educação Superior em parceria com a Universidade de Sorocaba (Uniso)	A revista foi publicada durante três anos, voltada a trabalhos relacionados à avaliação do Ensino Superior. Tratava-se de revista aberta à difusão, aos intercâmbios e aos debates de interesse da comunidade universitária, publicando prioritariamente trabalhos relacionados aos temas da avaliação da Educação Superior, em português ou em espanhol. Seus exemplares continuam on-line, apesar de a revista não estar sendo mais editada.	
Revista Meta: Avaliação	Revista eletrônica quadrimestral	http://metaavaliacao.cesgranrio.org.br/index.php/metaavaliacao/about/editorialPolicies#focusAndScope	Cesgranrio	“A revista do mestrado profissional em avaliação da Fundação Cesgranrio, <i>Meta: Avaliação</i> é uma publicação científica, de periodicidade quadrimestral, que veicula exclusivamente por meio da internet, nos meses de abril, agosto e dezembro. A publicação possui caráter multidisciplinar, porém com ênfase especial nas áreas de Educação, saúde e sociedade. <i>Meta: Avaliação</i> é um canal de divulgação da prática e da teoria da avaliação no Brasil e no mundo.” O periódico dedica-se a publicar pesquisas, resenhas de livros, teses e dissertações.	
Revista do Serviço Público	Revista trimestral	http://www.enap.gov.br/index.php?option=content&task=view&id=257	Escola Nacional de Administração Pública (Enap)	Revista voltada para a administração pública e a capacitação de gestores públicos. “A <i>Revista do Serviço Público</i> , editada pela Escola Nacional de Administração Pública há mais de 70 anos, publica artigos inéditos sobre temas relacionados a Estado e sociedade; políticas públicas e administração pública. Os artigos passam por análise prévia de adequação pela editoria da revista e posterior avaliação de dois a três pareceristas por sistema de <i>blind review</i> .” A Enap também enfoca a avaliação de políticas públicas.	
Revista de Gestão e Avaliação Educacional**	Revista eletrônica	http://cascavel.ufsm.br/revistas/ojs-2.2.2/index.php/regae/index	Universidade Federal de Santa Maria, Rio Grande do Sul (UFSM)	A <i>Revista de Gestão e Avaliação Educacional (Regae)</i> publica trabalhos inéditos resultantes de pesquisas já concluídas, pesquisas com resultados preliminares e estudos teóricos sobre a temática da revista, ou seja, gestão e avaliação educacional.	
Revista Avaliação em Foco	Revista trimestral	http://www.avaliacaoenfoco.com.br/	Grupo Avaliação em Foco	<i>Avaliação em Foco</i> é uma publicação voltada aos professores e gestores de escolas e redes públicas de ensino, que coloca o tema avaliação no centro das discussões acerca da qualidade de ensino. Aberta a receber artigos, prevê publicar oito edições entre 2013 e 2014.	

* Último acesso em setembro de 2013.

** Revistas descontinuadas até o momento da publicação (setembro de 2013).

OPORTUNIDADES DE PUBLICAÇÃO – REVISTAS DIRECIONADAS À AVALIAÇÃO				
Nome	Tipo	Disponível em *	Organização	Descrição
American Journal of Evaluation (AJE)	Revista trimestral	http://www.eval.org/Publications/AJE.asp , http://aje.sagepub.com/		A AJE (<i>Revista Americana de Avaliação</i> , em português) publica artigos originais sobre métodos, teorias, práticas e conclusões de avaliações. O objetivo geral da publicação é apresentar o que há de melhor sobre o campo para aumentar a base de conhecimento e prática de seus leitores. Os artigos refletem a diversidade de tradições intelectuais, abordagens e práticas, além de domínios de aplicação. No entanto, é dada preferência aos textos que têm potencial de interessar uma vasta gama de avaliadores e que são escritos de forma acessível.
New Directions for Evaluation (NDE)	Revista trimestral	http://www.eval.org/Publications/NDE.asp , http://onlineilibrary.wiley.com/journal/10.1002/(ISSN)1534-875X	American Evaluation Association (AEA)	A revista NDE (<i>Novas Direções para a Avaliação</i> , em português), de periodicidade trimestral, é uma publicação oficial da Associação Americana de Avaliação que serve como fonte de dados. Publica trabalhos sobre todos os aspectos da avaliação, com ênfase na apresentação apropriada de reflexões, com questões de ponta sobre a teoria, a prática, os métodos e a profissão da avaliação, além do contexto organizacional, cultural e da sociedade em que ela ocorre. Cada edição é destinada a um único tópico, com contribuições encomendadas, organizadas, editadas e revisadas por um ou mais editores convidados.
Evaluation: The International Journal of Theory, Research and Practice	Revista trimestral	http://evi.sagepub.com/	Vinculada à European Evaluation Society (EES), The Tavistock Institute e publicada pela Sage	A revista <i>Evaluation</i> (Avaliação) tem como objetivo avançar na teoria, metodologia e prática da avaliação, juntando conhecimento e expertise de muitas disciplinas, profissões e países. Sua intenção é ser um fórum internacional de diálogo e debate entre acadêmicos e profissionais. Foi criada em 1995, pertencendo à SAGE, uma das maiores editoras acadêmicas do mundo.
Canadian Journal of Program Evaluation	Revista quadrimestral	http://www.evaluationcanada.ca/site.cgi?s=4&ss=2&lang=EM	Canadian Evaluation Society (CES/CSE)	A <i>Revista Canadense de Avaliação de Programa</i> é publicada três vezes ao ano pela CES (Sociedade Canadense de Avaliação, em português). Ela busca promover a teoria e a prática de avaliações de programas no Canadá, publicando artigos sobre todos os aspectos da teoria e prática da avaliação – incluindo metodologias, parâmetros de avaliação, implementação de avaliações, relatórios e uso de estudos e auditoria, ou “meta-avaliação”, da avaliação. A revista também publica artigos sobre os desafios da prática de avaliação, apresentando casos reais para a avaliação de profissionais; notas sobre a prática da avaliação, sobre todos os aspectos da prática da avaliação, que têm por objetivo compartilhar conhecimento prático, experiências e lições aprendidas para o benefício da comunidade de avaliadores; e resenhas de livros que trazem críticas a publicações de interesse e relevância no campo (compilações e autorais).

* Último acesso em setembro de 2013.

OPORTUNIDADES DE PUBLICAÇÃO – REVISTAS DIRECIONADAS À AVALIAÇÃO					
Nome	Tipo	Disponível em *	Organização	Descrição	
Evaluation Review	Revista bimestral	http://erx.sagepub.com/	Sage	A revista <i>Evaluation Review</i> (Resenha da Avaliação, em português) é “um fórum interdisciplinar para pesquisadores da área de ciências sociais, profissionais de planejamento e criadores de políticas que desenvolvem, implementam e usam estudos designados a melhorar a condição humana”. Publicada há três décadas sob responsabilidade de diferentes editores e publicada pela Sage.	
Journal of Development Effectiveness	Revista trimestral	http://www.3ieimpact.org/en/publications/journal/	International Initiative for Impact Evaluation (3ie)	A <i>Revista de Eficiência do Desenvolvimento</i> (tradução do título para o português) publica artigos relatando evidências do impacto de intervenções para o desenvolvimento. A revista não comunga de nenhuma abordagem de avaliação de impacto, mas exige que as técnicas usadas sejam aplicadas rigorosamente, dando preferência a estudos que sejam bem contextualizados e com uso apropriado de diferentes métodos. Ela também publica artigos de natureza mais conceitual referentes à avaliação de impacto, além de trabalhos que tratem dos aspectos práticos da condução de estudos de impacto. A revista tem uma política explícita de cobrir os aspectos práticos da condução de estudos de impacto. O periódico tem ainda a política explícita de “aprender com nossos erros”, desencorajando a publicação de artigos que tendem a apresentar resultados positivos. Por outro lado, artigos relatando intervenções com nenhum impacto, ou impacto negativo, são bem-vindos. Uma lista sobre novos estudos de impacto é incluída a cada edição.	
Journal of Multidisciplinary Evaluation	Revista eletrônica semestral	http://journals.sfu.ca/jmde/index.php/jmde_1/index	The Evaluation Center/Western Michigan University, Estados Unidos	Journal de acesso aberto. Publica artigos sobre avaliação de forma ampla: editoriais, introdução, artigos, revisão de livros, “Ideas to Consider”, entre outros.	

* Último acesso em setembro de 2013.

OPORTUNIDADES DE PUBLICAÇÃO – REVISTAS ACADÊMICAS (ÁREAS DE SAÚDE E EDUCAÇÃO)				
Nome	Tipo	Disponível em*	Organização	Descrição
Cadernos Cedes	Revista sem periodicidade definida	http://www.cedes.unicamp.br/cad_apresentacao.htm	Centro de Estudos Educação e Sociedade da Universidade Estadual de Campinas, São Paulo (Cedes – Unicamp)	<i>Cadernos Cedes</i> são publicações periódicas do Centro de Estudos Educação e Sociedade (Cedes), de caráter temático, dirigidas a profissionais e pesquisadores da área educacional com o propósito de abordar questões que se colocam como atuais e significativas nesse campo de atuação. Os <i>Cadernos</i> são editados desde 1980, mantendo de três a seis novos títulos por ano.
Revista Educação & Sociedade	Revista quadrimestral	http://www.cedes.unicamp.br/rev_apresentacao.htm	Cedes – Unicamp	Idealizada como instrumento de incentivo à pesquisa acadêmica e debate amplo sobre o ensino nos seus diversos prismas, a publicação atinge, após anos de publicação ininterrupta, um grande acúmulo de análises, informações, debates, fontes teóricas, relatos de experiências pedagógicas, entre outros, de grande interesse para cientistas e educadores que atuam nas ciências humanas e outros campos do saber.
Cadernos de Saúde Pública	Revista mensal	http://www4.ensp.fiocruz.br/csp/	Escola Nacional de Saúde Pública Sergio Arouca – Fundação Oswaldo Cruz	Publica artigos originais que contribuem para o estudo da saúde pública em geral e disciplinas afins, como epidemiologia, nutrição, parasitologia, ecologia e controle de vetores, saúde ambiental, políticas públicas e planejamento em saúde, ciências sociais aplicadas à saúde, entre outras.
Ciência & Saúde Coletiva	Revista bimestral	http://www.abrasco.org.br/publicacoes/csc.php	Associação Brasileira de Saúde Coletiva (Abrasco)	Publica debates, análises e resultados de investigações sobre um tema específico considerado relevante para a saúde coletiva.
Lua Nova	Revista quadrimestral	http://cedec.org.br/luanova.asp?In=current&page=fIn&subpage=rIn	Centro de Estudos de Cultura Contemporânea (Cedec)	Fundada em 1984, como uma revista de debate e de intervenção nas grandes questões que mobilizavam a sociedade na época. A partir de 1988, com o número 15, ganhou seu formato atual, com números centrados em núcleos temáticos e com artigos avulsos voltados para a reflexão de longo alcance, tanto na análise empírica quanto nos fundamentos teóricos, sobre questões relativas a três campos básicos: a democracia, a cidadania e os direitos. Atualmente, é publicação de primeira linha no campo acadêmico, com altas avaliações nos órgãos de fomento à pesquisa e com indexação nacional e internacional. Dedicar-se a fomentar a reflexão que ofereça fundamentos ao debate em profundidade de questões substantivas da atualidade, com ênfase na análise de políticas públicas buscando nelas o que têm de público e do panorama internacional com ênfase na posição brasileira. Dedicar também especial atenção à reflexão teórica nacional e internacional (na qual publica textos dos principais autores). Orienta-se também de modo sistemático para a publicação de trabalhos de autores jovens ou iniciantes.

* Último acesso em setembro de 2013.

OPORTUNIDADES DE PUBLICAÇÃO – REVISTAS ACADÊMICAS (ÁREAS DE SAÚDE E EDUCAÇÃO)					
Nome	Tipo	Disponível em*	Organização	Descrição	
Novos Estudos – Cebrap	Revista quadrimestral	http://www.novosestudos.com.br/vi/	Centro Brasileiro de Análise e Planejamento (Cebrap)	Publica estudos relevantes sobre temas das ciências sociais e humanidades, contribuindo para o debate de ideias em âmbito nacional e internacional, com discussões em áreas diversas, desde crítica de arte até implementação de políticas econômicas e sociais.	
Physis: Revista de Saúde Coletiva	Revista trimestral	http://www.ims.uerj.br/i_publicacao_physis.php	Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro (IMS - UERJ)	Sua orientação editorial é aberta para a produção na área de saúde coletiva, com ênfase nas áreas de ciências humanas e sociais e de política, planejamento e administração em saúde. Dada a existência de publicações relevantes nacionais na área de epidemiologia, o corpo editorial optou por priorizar artigos que abordem dimensões conceituais da área ou ainda repercussões sociais dos conhecimentos produzidos pela pesquisa epidemiológica. A revista é composta por artigos de livre submissão dos autores e possui uma seção de resenhas e críticas bibliográficas, uma seção de entrevistas e uma de cartas. Ocasionalmente, são propostos temas específicos considerados relevantes, com textos de autores convidados, com inserção relevante no debate sobre o tema. Mesmo esses, contudo, são submetidos à revisão por pares pelos editores da publicação.	
Série Estudos em Saúde Coletiva (Sesc)	Revista eletrônica	http://www.ims.uerj.br/i_publicacao_serie_coletiva.php	IMS UERJ	Publicação voltada à difusão de textos orientados para ampliar o debate intelectual no campo da saúde coletiva. Nesse sentido, é um veículo de apresentação da produção corrente do IMS - UERJ, no qual continua sendo prioritária a publicação de textos de autoria de professores e alunos.	
Revista de Administração	Revista trimestral	http://www.rausp.usp.br/	Departamento de Administração da Faculdade de Economia, Administração e Contabilidade – Faculdade de Economia e Administração da Universidade de São Paulo (FEA-USP)	A revista publica artigos de autores nacionais e estrangeiros selecionados com base em critérios de originalidade, qualidade e criatividade.	

* Último acesso em setembro de 2013.

OPORTUNIDADES DE PUBLICAÇÃO – REVISTAS ACADÊMICAS (ÁREAS DE SAÚDE E EDUCAÇÃO)				
Nome	Tipo	Disponível em*	Organização	Descrição
Revista de Gestão (Rege)	Revista trimestral	http://www.regeusp.com.br/	Departamento de Administração da Faculdade de Economia, Administração e Contabilidade (FEA-USP)	Publicação trimestral para divulgação de trabalhos de natureza acadêmica relacionados a temas da administração de empresas. Seu objetivo é colaborar para a disseminação do conhecimento por meio da publicação de trabalhos que atendam a critérios de qualidade científica e relevância na área. Busca ainda obter diversidade e abrangência nos temas publicados, bem como na origem geográfica dos autores, de modo a garantir uma cobertura ampla da pesquisa realizada em âmbito nacional. Procura também incentivar o envio de artigos em inglês ou espanhol, com o objetivo de permitir a disseminação de seu conteúdo internacionalmente.
Revista de Administração Pública	Revista bimestral	http://ebape.fgv.br/publicacoes/rap	Fundação Getúlio Vargas (FGV)	Publicação aberta às contribuições de todas as áreas de conhecimento preocupadas com a efetividade e equidade da ação pública, entre as quais se destacam: administração e desenvolvimento; administração pública; ciências humanas, sociais e sociais aplicadas; estudos organizacionais; gestão social; movimentos sociais; política pública. Segue, em seus princípios, a missão da FGV: avançar nas fronteiras do conhecimento na área das ciências sociais e afins, produzindo e transmitindo ideias, dados e informações, além de conservá-los e sistematizá-los, de modo a contribuir para o desenvolvimento socioeconômico do País, a melhoria dos padrões éticos nacionais, uma governança responsável e compartilhada e para a inserção do Brasil no cenário internacional.
Revista de Saúde Pública	Revista bimestral	http://www.rsp.fsp.usp.br/mensagem/pub/bemvindo.php?tipo=0	Faculdade de Saúde Pública da USP	Tem por finalidade publicar contribuições científicas originais sobre temas relevantes para a saúde pública em geral.
Saúde e Sociedade	Revista trimestral	http://www.scielo.br/scielo.php?script=sci_serial&pid=0104-1290&lng=pt&nrm=isso	Faculdade de Saúde Pública da USP e Associação Paulista de Saúde Pública	A publicação divulga a produção das diferentes áreas do saber sobre práticas da saúde, visando ao desenvolvimento interdisciplinar do campo da saúde pública. Destina-se à comunidade de profissionais, docentes, pesquisadores, especialistas da área de saúde pública/coletiva e de disciplinas afins. Desde sua criação, tem publicado trabalhos de diferentes campos que se relacionam ou que tenham como objeto de preocupação a saúde pública/coletiva. Nesse sentido, abarca a produção de diferentes ramos das ciências humanas e sociais e da emergente ciência ambiental, incorporando a produção científica, teórica e aquela mais especificamente relacionada às propostas de intervenção e prática institucional.

* Último acesso em setembro de 2013.

OPORTUNIDADES DE ARTICULAÇÃO				
Nome	Tipo	Disponível em*	Organização	Descrição
Rede Brasileira de Monitoramento e Avaliação	Rede de avaliadores e interessados	http://redebrasileira.demea.ning.com/	Diversas. Apoio: Fundação João Pinheiro (FJP), Banco Interamericano de Desenvolvimento, Banco Mundial	Seção da Rede de Monitoramento e Avaliação da América Latina e Caribe, fundada em novembro de 2008, com o apoio da Fundação João Pinheiro (FJP), do Banco Interamericano de Desenvolvimento e do Banco Mundial. Os principais objetivos são: promover o uso de práticas de monitoramento e avaliação; promover a troca de experiência e de instrumentos de trabalho; facilitar a identificação das instituições trabalhando nesse tema no Brasil; formar uma "comunidade de prática".
Projeto Avaliação	Grupo voltado à avaliação	http://institutofonte.org.br/projeto-avaliacao	Instituto Fonte	Iniciativa realizada pelo Instituto Fonte com apoio da Fundação Itaú Social e com diversas parcerias com Fundação Roberto Marinho, Instituto Paulo Montenegro, Ibope Inteligência, entre outros. O objetivo é o "fortalecimento do campo da avaliação de programas e projetos Sociais no Brasil por meio de estudos, articulação e espaços de formação". Realiza encontros abertos e gratuitos para a discussão de temas relacionados à avaliação. Também realiza pesquisas sobre o campo e a sistematização de casos de avaliação de iniciativas sociais. O material do projeto é disponibilizado no site.
Associação Brasileira de Avaliação Educacional (Abave)	Associação voltada à avaliação	http://www.abave.org.br/		Associação de profissionais interessados no tema da avaliação educacional. Realiza encontros e publicações, não tem uma revista periódica. "A Abave é um espaço plural e democrático para o intercâmbio de experiências entre os acadêmicos e os implementadores da avaliação educacional. O que justifica a sua criação é o crescimento, nos últimos anos, do número de pessoas que lidam com a avaliação e que começam a vê-la como seu campo de atuação profissional, seja como pesquisadores ou como responsáveis pela condução das políticas de avaliação nos sistemas educacionais. Mais importante ainda, observa-se a disposição desses dois grupos de se relacionarem entre si para aprender e maximizar o potencial dos seus trabalhos a favor da melhoria da Educação."
International Organization for Cooperation in Evaluation (IOCE)	Grupo voltado à avaliação	http://www.ioce.net/		Organização internacional que congrega 24 associações e redes de avaliação no mundo. "A Organização Internacional para a Cooperação na Avaliação (IOCE, na sigla em inglês) é uma aliança informal de organizações avaliadoras regionais e nacionais (associações, sociedades e redes) de todo o mundo que colaboram para: construir liderança e capacidades em avaliação em países em desenvolvimento; promover a troca de teorias e práticas de avaliação mundo afora; lidar com os desafios internacionais da avaliação e ajudar os profissionais a usar uma abordagem mais global, contribuindo para a identificação e solução de problemas."

* Último acesso em setembro de 2013.

OPORTUNIDADES DE ARTICULAÇÃO						
Nome	Tipo	Disponível em *	Organização	Descrição		
Clear Initiative	Organização de suporte à avaliação	http://www.theclearinitiative.org/index.html		Ligada ao Banco Mundial, a Clear dá suporte ao desenvolvimento de grupos que fortalecem a avaliação regionalmente, especialmente por meio da capacitação e produção de publicações. "A Clear é um esforço colaborativo de doadores e países parceiros com o objetivo de fortalecer os sistemas e capacidades de monitoramento e avaliação das nações participantes, além da gestão de desempenho para embasar o foco nos resultados."		
International Initiative for Impact Evaluation (3ie)	Organização de suporte à avaliação	http://www.3ieimpact.org/		Organização que financia estudos e iniciativas de avaliação de impacto, financiada pela Fundação Bill e Melinda Gates e outras. Os objetivos da instituição são: "gerar novas evidências do que funciona, sintetizar e disseminar essas evidências, construir uma cultura de produção de políticas a partir de evidências e desenvolver a capacidade e o uso das avaliações de impacto".		
European Evaluation Society (EES)	Associação voltada à avaliação	http://www.europeanevaluation.org/		O objetivo da EES (Sociedade Europeia de Avaliação, em português) é estimular e promover a teoria, a prática e a utilização da avaliação de alta qualidade na Europa, em particular, mas não exclusivamente. Esse objetivo é alcançado com a reunião de acadêmicos e profissionais de todo o continente, de diferentes setores, criando um fórum em que todos os participantes podem se beneficiar da cooperação e oportunidades de colaboração. O principal veículo para essa troca de conhecimento é a conferência bienal. Além disso, a EES trabalha para oferecer uma plataforma de interação, discussão e contato entre os profissionais de avaliação, por meio de ferramentas disponíveis na internet, fóruns e eventos.		
American Evaluation Association (AEA)	Associação voltada à avaliação	http://www.eval.org/		Associação de referência em avaliação no mundo, promove encontros anuais e tem entre seus objetivos promover propósitos científicos e educacionais, melhorar a teoria, prática e métodos da avaliação; aumentar o uso da avaliação; promover a avaliação como profissão; e apoiar a contribuição da avaliação para a geração de teoria e conhecimento sobre a ação humana eficiente.		
Canadian Evaluation Society	Associação voltada à avaliação	http://evaluationcanada.ca/		Organização dedicada ao avanço da prática e teoria da avaliação. No campo da avaliação, promove: liderança, conhecimento, suporte ativo e desenvolvimento profissional.		
Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe (ReLAC)	Rede de avaliadores e interessados	http://www.relacweb.org		Conjunto de redes voltado a contribuir com o fortalecimento de capacitação em monitoramento e avaliação, além de profissionalizar a função de avaliação na América Latina.		

* Último acesso em setembro de 2013.

OPORTUNIDADES DE ARTICULAÇÃO				
Nome	Tipo	Disponível em*	Organização	Descrição
Red de Monitoreo y Evaluación de América Latina y el Caribe (REDLACME)	Rede de avaliadores e interessados	http://redlacme.org/		A Rede de Monitoramento e Avaliação (M&A), apoiada desde 1995 pelo Banco Mundial, pelo Programa Prodev do Banco Interamericano de Desenvolvimento (BID) e por um crescente grupo de países da América Latina. Pretende servir de cenário para a análise de resultados e de dificuldades encontrados durante a institucionalização dos sistemas M&A nos níveis nacionais e subnacionais, bem como promover o diálogo, compartilhar experiências, fortalecer a capacidade institucional e estimular a colaboração.
Evaluation Cooperation Group (ECG)	Grupo voltado à avaliação	https://wppq4.adb.org/LotusQuic/kr/ecg/Main.nsf/h_Toc/c4d2972a55a9514948257bbe0023aa84/?OpenDocument	Grupo dos Bancos Multilaterais de Desenvolvimento	O ECG (Grupo de Cooperação de Avaliação, em português) foi criado pelos chefes de avaliação de bancos de desenvolvimento multilaterais (MBDs, na sigla em inglês), em 1996, para fortalecer o uso da avaliação e contribuir para sua disseminação, para obter maior eficiência e responsabilidade nos serviços de avaliação dessas instituições, harmonizar os indicadores de desempenho, metodologias e abordagens de avaliação, melhorar o profissionalismo da avaliação internamente e a colaboração entre diretores de avaliação de organizações bilaterais e multilaterais de desenvolvimento e facilitar o envolvimento de países membros devedores na avaliação, além de construir sua capacitação na área.
The Evaluation Capacity Development Group – ECDG	Grupo voltado à avaliação	http://www.ecdg.net/	ECDG	O ECDG (Grupo para o Desenvolvimento da Capacidade Avaliativa, em português) é uma organização norte-americana, fundada por profissionais interessados na promoção da cultura de avaliação, pertencentes a diversos países. Disponibiliza textos, oportunidades de formação e possibilidade de debate entre profissionais.
DAC Network on Development Evaluation	Grupo voltado à avaliação	http://www.oecd.org/dac/evaluation/	OCDE	Rede da OCDE (sigla em inglês para Organização para a Cooperação e o Desenvolvimento Econômico) voltada para o tema da avaliação do Desenvolvimento e para contribuir para promover “melhores resultados do desenvolvimento utilizando avaliação para construção de bases com fortes evidências para a elaboração de políticas e para a aprendizagem”.
Independent Evaluation Group	Grupo voltado à avaliação	http://ieg.worldbankgroup.org/	Banco Mundial	Grupo orientado à avaliação das iniciativas do Banco Mundial em diversas esferas. Pública parte das avaliações que realiza.
The United Nations Evaluation Group (UNEG)	Grupo voltado à avaliação	http://www.uneval.org/	Organização das Nações Unidas (ONU)	Grupo de unidades profissionais que atuam com avaliação dentro do sistema ONU. “O UNEG (Grupo de Avaliação das Nações Unidas, em português) é uma rede profissional que junta unidades responsáveis por avaliação no sistema da organização, incluindo agências especializadas, fundos, programas e entidades afiliadas [...] Busca aumentar a objetividade, a eficiência e a visibilidade da função de avaliação no sistema da ONU e defender a importância da avaliação para o aprendizado, tomadas de decisões e responsabilidade.” Também publica editais de contratação de serviços de avaliação para a ONU.

* Último acesso em setembro de 2013.

OPORTUNIDADES DE ARTICULAÇÃO					
Nome	Tipo	Disponível em *	Organização	Descrição	
Unicef Evaluation Office	Grupo voltado à avaliação	http://www.unicef.org/evaluation/index_60801.html	Fundo das Nações Unidas para a Infância (Unicef)	Agrega e articula as avaliações realizadas dos projetos vinculados ao Unicef. “O Unicef conduz avaliações em cinco níveis institucionais, refletindo o sistema de responsabilidade organizacional: local ou projeto; programa nacional de cooperação; regional; estratégico global; e desempenho institucional. A descentralização é característica-chave do sistema de avaliação do Unicef, junto com países do programa, escritórios nacionais da entidade vão continuar a encomendar a maior parte dos trabalhos de avaliação. A avaliação faz três perguntas-chave: Estamos fazendo a coisa certa? Estamos fazendo do modo certo? Como podemos fazer melhor?”.	
International Organization for Collaborative Outcome Management (IOCOM)	Plataforma de partilha de informações e e-learning	http://www.iocomsa.org/node/1		A IOCOM é uma organização virtual de profissionais e acadêmicos, e uma aliança de organizações nacionais e internacionais (associações, sociedades e redes) comprometidas com a disciplina da gestão de resultado e desenvolvimento. Criou um fórum on-line de troca de teorias, metodologias e práticas e oferece conexões globais para profissionais de gestão de resultado e desenvolvimento, além de notícias de eventos e iniciativas.	
Desarrollo, Evaluación & Complejidad	Blog	http://www.scoop.it/t/albordedelcaos		Blog com curadoria de Pablo Rodríguez-Bilella, sociólogo e avaliador argentino, membro do Comitê Executivo da Ideas (International Development Evaluation Association, ver “oportunidades de formação”), consultor na Conicet (Conselho Nacional de Investigações Científicas e Técnicas do governo argentino) e professor na Universidad Nacional de San Juan. No blog, Pablo disponibiliza notícias e links relacionados com avaliação.	
Evaluación Políticas Públicas	Grupo voltado à avaliação	http://www.evaluacionpoliticaspublicas.com/evaluoteca/	Núcleo de Avaliação de Políticas Públicas da Universidade do Chile (NEPP, em espanhol)	Site do Núcleo de Avaliação de Políticas Públicas da Universidade do Chile, disponibiliza oportunidades, textos e uma base de documentos produzidos pelo próprio Núcleo.	
Evaluatca – Información, novedades, opiniones y comentarios críticos sobre Evaluación	Blog	http://evaluoteca.wordpress.com/		Blog com curadoria de Rafael Monterde Díaz, professor na área de gestão de projetos e diretor de programa de pós-graduação nesse tema na Universidade Politécnica de Valência, Espanha. Disponibiliza notícias, textos, oportunidades de troca entre profissionais.	

* Último acesso em setembro de 2013.

Agradecimentos

Um grande grupo de profissionais participou ativamente da organização do II Seminário Internacional “Avaliação para o Investimento Social Privado: Metodologias” e foi fundamental para o alcance de seus objetivos. Da concepção do evento à produção deste livro, queremos registrar nossos mais sinceros agradecimentos a:

Aline Chagas
Ana Carolina Moraes
Andre Degenszajn
Anna Lívia Solon Arida
Antônio Ribeiro
Carlos Eduardo Garrido
Cláudio Silva
Daniel Brandão
Donna M. Mertens
Geraldine de Araújo Silva
Madelene Barbosa
Martina Rillo Otero
Miguel Székely
Priscila Dias Leite
Thomaz Kauark Chianca
Vanessa Orban
Vanessa Panchieri

Este livro foi composto nas fontes Milo e Milo Serif e impresso em outubro de 2013
pela Corprint Gráfica e Editora Ltda., sobre papel offset 90g/m².

ORGANIZAÇÃO



PARCERIA



avaliação e estratégia em desenvolvimento social

APOIO



Fundação Santillana

ISBN 978-85-63489-16-6



9 788563 489166

“Os métodos dialéticos [de avaliação] engajam, em um diálogo significativo, atores que aparentemente detêm posições díspares, usando múltiplas estratégias dialógicas e de negociação. Em sua linha investigativa, podem incluir observação, experiência e experimentos, com a expectativa de que dados empíricos múltiplos não vão convergir tipicamente em uma conclusão simples, mas que fornecerão divergências que resultarão em melhor compreensão do objeto...”

DONNA M. MERTENS

“...isolar a causalidade entre determinada ação e o efeito sobre um indivíduo ou grupo é tarefa difícil. Os programas são aplicados em um contexto no qual uma ampla gama de forças e elementos interagem simultaneamente, e a capacidade de extrair uma resposta precisa de uma política específica também pode ser virtualmente impossível.”

MIGUEL SZÉKELY

“... a avaliação vai muito além da aplicação de métodos de pesquisa para estudar variáveis relacionadas ao “avaliando”, normalmente as que são de medição direta, rápida e confiável. O que muitos profissionais [...] não se perguntam [...] é se estas variáveis são os aspectos realmente importantes a serem considerados [...]”

THOMAZ KAUARK CHIANCA