

# Avaliações sob Medida – Produzir Estudos Relevantes em Serviços de Saúde Reais

Rogério Renato Silva

Max Felipe Vianna Gasparini

Madelene Barboza

## Introdução

A avaliação em saúde constitui um instrumento técnico-político essencial para que os direitos do cidadão sejam respeitados e os serviços sejam acessíveis e de qualidade. Se o cotidiano da gestão da saúde requer de seus agentes a habilidade de administrar a pressão por mais atendimentos e menores custos, mais tecnologia e humanização, resolutividade e promoção, e se é crescente a expectativa por serviços com maior capacidade estratégica, eficiência, transparência e controle social, há sinais inquestionáveis de que são necessárias transformações, e as práticas de avaliação podem auxiliar nesse sentido.

Contudo, a história revela quão paralisantes e desmobilizadoras podem ser as avaliações. Orientadas mais ao controle do que à aprendizagem organizacional ou submetidas a processos excessivamente burocráticos, certas avaliações não conseguem ir além da coleta sistemática e obsessiva de informações irrelevantes, representando um desperdício de tempo e energia, preciosos para equipes, gestores e usuários. Assim, de que modo as armadilhas podem ser evitadas e as avaliações, potencializadas?

Neste capítulo, abordamos maneiras de pensar e realizar avaliações que se conectem de forma dinâmica ao cotidiano das organizações de saúde. Em quais premissas os processos de avaliação podem se apoiar para fazerem sentido no cotidiano dos serviços? Como abordar os objetos

avaliativos de forma sistêmica? Como construir avaliações que permitam formar juízos de valor justos sobre a realidade? De que maneira articular avaliação e aprendizagem? Quais estratégias podem favorecer uma cultura de avaliação que relacione investigação, descoberta e decisão? São perguntas como essas que buscamos responder nas próximas páginas.

Para isso, associamos, em uma narrativa informal, reflexões teóricas, memórias da prática, armadilhas técnicas e aconselhamento político. Nossa produção foi escrita na perspectiva de autores há anos imersos em avaliações de distintos objetos do campo social, trabalhando tanto no âmbito público como no privado. Nosso intuito é que o texto chegue aos leitores como a reafirmação de que não haverá sistema de saúde universal, integral e equânime sem instituições democráticas que tenham capacidade de compreender sua relevância e mérito e preservem a disposição genuína para reinventar-se e fazer sempre melhor.

## O que governa uma avaliação?

Em nosso meio, existe a ideia de que as avaliações têm vida própria. Assim como as jabuticabas, as avaliações seriam frutos de árvores frondosas, chegando ao mundo com certa sazonalidade para o deleite de seus consumidores. A imagem, em alguma medida engraçada, se assenta na perigosa crença de que as avaliações seriam função automática ou natural da gestão, sendo governadas por certa racionalidade técnica inerente às organizações e independente dos atores e da política organizacional. Se a imagem da jabuticaba não deixa de ser engraçada, ela é também estapafúrdia.

Toda avaliação é governada por interesses no seio de uma organização. Toda avaliação está nas mãos de um ator particular. Toda avaliação responde a certo vetor de poder, podendo satisfazer a um ator enquanto deixa insatisfeito ou mesmo submete outro. Na medida em que todo processo avaliativo está imerso em uma cultura organizacional específica, toda avaliação sofre a influência da dinâmica dessa organização, dos atores que a compõem e dirigem, das forças internas e externas que a sustentam, das disputas que a modelam e dos interesses técnicos e econômicos, éticos e políticos que nela imprimem suas marcas. Ao contrário das jabuticabas, nada há de natural nesses cenários, nos quais as avaliações sempre têm nome e sobrenome.

Por isso, identificar os atores-chave e perceber como se dá o governo das avaliações constituem ações estratégicas. No entanto, tendo em vista que essas condutas costumam ser negligenciadas, constatamos que o processo de disputa, pelo governo, das avaliações se dá frequentemente com falhas. Sem atenção e sem identificar agendas, o processo ora despolitiza as escolhas para fazer valer um discurso técnico de neutralidade científica, ora submete a avaliação a lógicas autoritárias, nas quais apenas um ator governa a todos e por todos. Como, então, evitar a armadilha?

Começamos pensando que toda avaliação diz respeito a um objeto que pertence a diferentes sujeitos, ou seja, um objeto ao qual diversas pessoas estão vinculadas. Tomemos como exemplo um programa municipal de combate à dengue. Ele pertence à secretaria de saúde tanto quanto à população de um território. Pertence aos técnicos da vigilância tanto quanto aos profissionais de um pronto atendimento que lidam com os quadros agudos da doença. Mesmo admitindo-se que haja diferentes níveis de vínculo e propriedade sobre o processo, que haja papéis muito distintos para os atores, não restam dúvidas de que tal programa é um objeto que pertence a esses sujeitos, um objeto produzido por suas ações técnicas e políticas.

É o conjunto de vínculos a esse objeto que sustenta que o governo da avaliação do programa de combate à dengue deveria se dar de forma que essa avaliação, com sua direção política e desenho metodológico, fosse uma construção desses atores. Primeiro, porque reconhecemos que um programa dessa natureza se inscreve no Sistema Único de Saúde (SUS) e, portanto, submetido a seus princípios, como o controle social. Depois, porque a avaliação de um programa como esse deveria estar comprometida em ampliar a efetividade do combate à dengue no município, o que fatalmente implicaria mais responsabilidades para seus atores. Uma vez que que aspiramos a mais corresponsabilidade, compromisso e um melhor aproveitamento dos recursos públicos, não podemos separar a avaliação do processo político que dirige a ação; não podemos cindir avaliação e política: ao contrário, queremos fazer política com avaliação.

É importante observar que a advertência sobre o componente político das avaliações é consequência da abertura teórica propiciada pela crítica aos modelos avaliativos tradicionais. Autores como Guba e Lincoln (2011) ocupam posição de destaque nesse processo, notadamente em virtude de suas construções sobre a avaliação de quarta geração ou avaliação construtivista. Suas propostas são baseadas em críticas epistemológicas, metodológicas e políticas dos modelos avaliativos pautados predominantemente pelo paradigma positivista da realidade, o que levou à ideia de avaliações isentas de valores e posicionamentos políticos, liberando os avaliadores de quaisquer responsabilidades.<sup>1</sup>

Nessa mesma direção, House (2004) problematiza o papel dos avaliadores em um contexto no qual as premissas da avaliação estão cada vez mais submetidas aos interesses políticos. O autor chama a atenção para a necessidade de estar atento às pressões políticas que cercam as avaliações, ressaltando que a independência de um processo avaliativo passa por sua capacidade de identificar e manejar os interesses dos grupos envolvidos, o que é fundamental para produzir avaliações mais democráticas.<sup>2</sup> O autor fundamenta a Avaliação Democrática Deliberativa (*Deliberative Democratic Evaluation*) em três princípios:

- 1) A inclusão de todas as informações relevantes e opiniões dos *stakeholders*, seus valores e interesses na avaliação;
- 2) Extenso diálogo entre avaliadores e demais *stakeholders*, buscando o entendimento entre todos; e
- 3) Deliberação feita por todas as partes para ajudar a alcançar conclusões válidas.<sup>2</sup>

Tendo em vista tais premissas teóricas e nossa experiência na realização de processos avaliativos, existem diversas maneiras de mobilizar e trazer os atores para uma avaliação, o que nem sempre é uma tarefa fácil. Em alguns casos, a ausência do ator é exatamente sua estratégia para direcionar o estudo à não legitimidade. De qualquer forma, há diversas estratégias capazes de garantir o engajamento dos atores, todas elas bastante semelhantes àquelas utilizadas para mobilizar pessoas em processos políticos. A Figura 2.1 ilustra os motivos que observamos mais frequentemente nas estratégias de envolvimento. Todas eles são capazes de incomodar, despertar, sensibilizar e facilitar o engajamento dos atores em uma avaliação.

Observando a Figura 2.1, podemos perceber que quaisquer das estratégias dizem respeito a circunstâncias específicas dos atores e suas realidades. Seja qual for o caminho de mobilização e engajamento escolhido para uma avaliação, deve-se ter em conta que, quanto maiores o vínculo e o engajamento no processo avaliativo, maiores as chances de o estudo se tornar relevante. Cousins e Whitmore (1998),<sup>3</sup> também pesquisadores do campo da avaliação colaborativa, identificam três dimensões fundamentais para pensar a participação em processos avaliativos:



Figura 2.1 – Categorias mais potentes para produzir envolvimento de atores.

- 1) A diversidade de atores envolvidos nos vários grupos de interessados em um objeto de avaliação;
  - 2) O grau de profundidade da participação desses atores no processo avaliativo; e
  - 3) O grau de controle que tais atores terão sobre as decisões do processo avaliativo.
- Para os autores, a atenção a tais dimensões ajudaria a garantir um envolvimento capaz de superar as formas sem influência verdadeira. Outrossim, quanto mais poder tiverem os atores que governam uma avaliação, maior será sua relevância. Jamais prefira as jabuticabas.

## Os modelos teóricos: as expectativas e a realidade

Um dos grandes desafios que enfrentamos ao fazer uma avaliação é lidar com as diferentes maneiras como os atores, incluindo os próprios avaliadores, enxergam o objeto avaliativo. É comum que existam leituras muito díspares dos resultados esperados, das estratégias e recursos utilizados em um programa, dos produtos entregues e até mesmo do público-alvo prioritário. Quando comparamos as formas como um objeto é percebido pelos atores, não é rara a impressão de que eles estejam a enxergar objetos distintos ou até mesmo antagônicos em alguns aspectos.

Há diferentes explicações para essa variedade de interpretações. Algumas vezes, as diferenças de leitura são oriundas de posições ético-políticas diferentes. Quem está em posição central de poder tende a cultivar imagens diferentes daqueles que estão na periferia, por exemplo. Em outros momentos, as diferenças ficam evidentes em função de se estar ou não na posição de gestor ou gestora, o que fará a diretora de um programa ter leituras distintas daquelas de um médico de família, que está na outra ponta da cadeia. Em outras situações, é o acúmulo teórico o fator a diferenciar a percepção dos atores, antagonizando o acadêmico e o usuário, o profissional com ensino superior e o agente de saúde, entre tantos outros exemplos.

Contudo, a principal explicação para as diferenças de leitura diz respeito ao quanto os atores estão mergulhados no objeto avaliado e ao quanto os cotidianos da gestão e do cuidado capturam sua atenção para especificidades, impedindo-os de formar imagens mais abrangentes da realidade. Na medida em que conviver com leituras parciais, imprecisas ou mesmo

errôneas de um objeto distorce as expectativas e influencia as avaliações de forma negativa, é importante reconstruir os objetos de avaliação na perspectiva de um modelo teórico, de uma teoria da ação, retomando-os em seu contorno original, o que tem sido incentivado por diversos autores.

Tal reconstrução se faz necessária em virtude da preocupação em qualificar os processos avaliativos antes de neles investir. Os chamados estudos de avaliabilidade foram desenvolvidos inicialmente nos Estados Unidos, dos anos de 1970, e tinham como objetivo analisar se eventuais processos avaliativos auxiliariam a evitar avaliações desnecessárias ou impertinentes. Os estudos de avaliabilidade buscam esclarecer os objetivos de determinada intervenção, explorar a realidade na qual se desenvolvem suas atividades e identificar questões que permeiam as práticas dos grupos envolvidos na condução da iniciativa.<sup>4-6</sup> O campo da Saúde, no Brasil, vem utilizando as análises de avaliabilidade como políticas de redução de danos<sup>8</sup>, programas de promoção da saúde<sup>9</sup> e estratégia no âmbito da gestão do trabalho e provimento de profissionais.<sup>10</sup>

O uso de modelos teóricos parte do pressuposto de que toda ação programática em saúde tem uma teoria subjacente a suas operações, uma teoria da ação. Se toda ação é concebida e planejada em certo momento, podemos supor que ela surge em resposta a demandas e interesses específicos, consumindo recursos financeiros, materiais e humanos para entregar algo a alguém. O pressuposto em questão diz também que tal ação (e sua teoria) nem sempre é explícita para os atores, uma vez que pode ter sido concebida por outros atores ou advir de outro momento histórico. Percebemos que as imagens, com frequência, se perdem nas organizações, como se os planos originais perdessem a cor com o passar do tempo. É a lacuna criada por tais fenômenos que fundamenta a construção de modelos teóricos do objeto que pretendemos avaliar.

Certos autores, como Medina et al,<sup>7</sup> argumentam que qualquer avaliação precisa reconhecer seu objeto para representá-lo em seus componentes, fluxos, relações causais, pressupostos e expectativas. Em muitas circunstâncias, a reconstrução do modelo teórico traz elementos cruciais para analisar a relevância do objeto: sua adequação, arranjos técnicos, estratégias e tecnologias. Para os autores, a construção da teoria do objeto-modelo permite não apenas enxergar o programa que queremos avaliar, mas fazer escolhas técnicas e políticas durante uma avaliação.

Ainda segundo os autores, à medida que o modelo teórico explicita a aposta em um modelo de cuidado, abre-se a possibilidade de que uma avaliação compreenda em que medida tal aposta adere à realidade, que demandas efetivamente responde, que recursos mobiliza e, sobretudo, que resultados produz.<sup>7</sup> Se uma avaliação lida inevitavelmente com diferenças, discrepâncias e lacunas entre expectativas e realidade, entre o que era necessário e o que foi possível, é crucial que tais expectativas e necessidades estejam claramente definidas.

Nessa mesma linha, vem de Cassiolato e Guerese (2010) o argumento de que um modelo teórico implica o ordenamento dos componentes de um serviço, projeto ou programa deve ser feito de forma articulada, dando ênfase aos resultados a serem alcançados.<sup>11</sup> Mais do que reconhecer componentes, os autores argumentam que o uso de modelos teóricos esclarece as ideias (problemas, pressupostos) e as hipóteses (estratégias, lógica, resultados esperados) que conferem sentido e racionalidade a uma intervenção.

Segundo o Ministério do Desenvolvimento Social brasileiro (BRASIL, 2014),<sup>12</sup> um modelo teórico pode identificar as relações causais entre ações e resultados. Uma de suas características mais importantes é demonstrar, ao menos no plano ideativo, que a adequada articulação de certos recursos e profissionais pode significar a solução dos problemas que

desencadearam as ações, os programas ou os próprios serviços, o que não deixa de ser uma forma de projetar o impacto esperado na realidade.

Em Chen (1990), encontramos outra contribuição teórica importante a respeito dos modelos teóricos. O autor afirma que as teorias de avaliação têm característica prescritiva, ou seja, estão comprometidas não apenas a descrever e julgar a relevância e o mérito de um determinado objeto, mas em apontar o que deve ser feito para aprimorá-lo.<sup>13</sup> Para atingir essa finalidade, é essencial que qualquer avaliação lance mão de todo o conhecimento disponível sobre o objeto avaliado, envolvendo perspectivas, hipóteses e expectativas dos atores interessados no objeto e em sua avaliação.

Na perspectiva da abordagem integradora proposta por Chen (1990), a construção de uma teoria da ação subjacente a cada objeto avaliado, que podemos chamar de teoria do programa, deve incorporar tanto os saberes acadêmicos sobre o objeto avaliado como os saberes advindos de sua realidade política e operacional, o que converge com a proposição de Medina et al.<sup>7</sup> A abordagem integradora, somada à tradição democrática do SUS, nos permite propor a elaboração de modelos teóricos ou teorias de programa em diferentes etapas, que, segundo Cruz (2006), estão orientadas pela necessidade de produzir diálogos e acordos entre os atores envolvidos em uma avaliação.<sup>14</sup> Nesse ponto, compreendemos que o modelo teórico ganha em precisão e legitimidade à medida que é construído pelos atores que governam a avaliação. A Figura 2.2 resume essa ideia.

Um dos maiores desafios das organizações contemporâneas consiste em criar ambientes e cultura de diálogo, reflexão e construção coletiva das questões ligadas ao cotidiano do trabalho. Habitados a viver em caixas e a não atravessar fronteiras, muitos de nós carregamos dificuldades para o trabalho cooperativo e para a construção em grupo. Estejamos mais ou menos abertos a esse tipo de experiência, o fato é que as avaliações são bastante efetivas em demonstrar a importância de fazer circular a palavra entre os atores. Essa é a condição fundamental para um exercício que visa construir um modelo teórico a partir de palavras, histórias, lembranças e leituras.

Se “palavra que circula é saber que circula, poder que opera e se movimenta”,<sup>15</sup> a construção do modelo teórico para uma determinada intervenção que realmente faça sentido e se

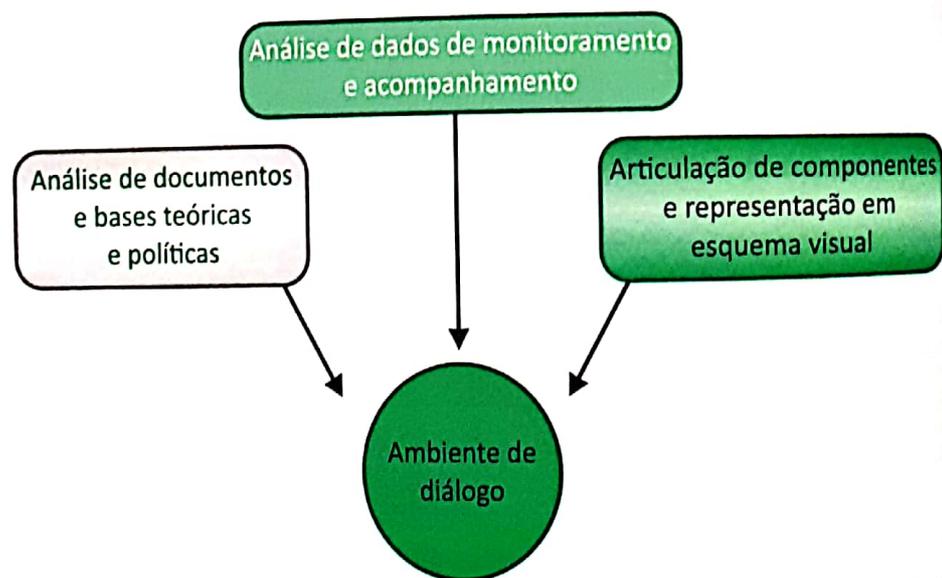


Figura 2.2 – Etapas para a elaboração de uma teoria da ação (teoria do programa).

apoie na experiência de todos os envolvidos não pode abrir mão de espaços onde a *palavra* possa circular sem constrangimentos ou coerções. É nesse sentido que, além do manejo cuidadoso do processo grupal, o uso de perguntas disparadoras ou norteadoras pode auxiliar a produção. As perguntas podem abrir janelas, identificar sentidos, clarear elos e ajudar a desenhar o modelo. O Quadro 2.1 mostra, como exemplo, uma sequência de perguntas capazes de apoiar o processo de construção.

### Quadro 2.1 – Perguntas-chave para a construção de modelos teóricos

1. Quais os principais problemas, agravos ou demandas de saúde que o serviço ou programa se propõe a combater?
2. Tomando como base tais problemas, agravos ou demandas, quais transformações de longo prazo ou que impactos o serviço ou programa pode alcançar?
3. Com quais atores o serviço ou programa trabalha? Quais estratégias são utilizadas para cada público? Como cada ator está sendo impactado por determinada estratégia?
4. Há resultados de curto e de longo prazo encadeados? Quais são eles?
5. Quais indicadores são ou têm sido capazes de mostrar o alcance de tais resultados de curto e médio prazo, bem como o impacto do serviço ou programa?
6. O caminho de transformação que o serviço ou programa percorre está claro? Há lacunas? Há pontos de maior potencial? Há condições necessárias que estão ou não estão acontecendo?
7. Em que medida a relação entre as estratégias e os resultados é consistente?
8. Há clareza sobre os resultados esperados?
9. Qual esquema visual melhor expressa o modelo de serviço ou do programa?

O exercício dialógico para a construção de um modelo teórico pode ganhar muita qualidade à medida que aos discursos e às novas construções somam-se a observação de atividades e a captura e análise de documentos organizacionais. Os projetos originais, termos de referência, relatórios de monitoramento, protocolos, prontuários e outras formas de registro da gestão e do cuidado podem representar um caudaloso oceano de memórias e saberes com potencial para revelar muito do que não é visível. Portanto, mesmo em uma etapa pré-avaliativa, ou seja, numa etapa em que o que se busca é a aproximação do objeto, muitas atividades de coleta e análise já podem ser feitas.

No exercício dialógico para a construção de um modelo teórico há aspectos-chave que merecem destaque, em razão da necessidade de receberem atenção especial dos atores, sobretudo daqueles que desempenham função de avaliadores:

- 1) O envolvimento de atores-chave que estejam de fato envolvidos com a avaliação e os saberes que ela produzirá;
- 2) O encontro da avaliação com seu objeto, a fim de que esse objeto seja percebido com a maior nitidez possível, crucial para o estudo;
- 3) A atenção ao conjunto de resultados prometido pelo programa ou serviço, essencial para a construção de critérios julgamento, como abordaremos mais adiante.

Se cabe uma última observação a respeito da construção dos modelos teóricos, ela diz respeito a quanto tem sido importante a utilização de esquemas visuais para dar tangibilidade ao modelo teórico. Os infográficos ou diagramas, como ilustra a Figura 2.3, devem ser peças capazes de resumir e evidenciar certa lógica, além de favorecer o diálogo entre os atores e facilitar a comunicação dos resultados do processo de trabalho. Há nos esquemas visuais bom potencial para demonstrar os componentes de um serviço ou programa, seu ordenamento, fluxo e encadeamento. Ante a complexidade dos objetos, as imagens são, muitas vezes, potentes dispositivos para disparar reflexão, auxiliando grupos a se surpreenderem com a realidade em curso.

Vejamos agora de que maneira o esforço empregado na construção do modelo teórico dá sustentação à etapa na qual o desenho técnico específico da avaliação será construído. Abordaremos a centralidade das perguntas avaliativas na condução de um estudo para, depois, ingressarmos na matriz avaliativa, dando ênfase aos indicadores e critérios para produzir juízo de valor (julgamento).

## As perguntas avaliativas para nortear os estudos

Ainda que a teoria de programa do objeto de avaliação seja consistente, explícita e reconhecida pelos atores, não será capaz de guiar a avaliação. O que muitos autores defendem e a prática avaliativa demonstra é que a formulação de boas perguntas é imprescindível para a produção de avaliações relevantes. Em sentido geral, uma pergunta resume o ato de interrogar a realidade e representa a necessidade de descoberta, o desejo de investigação, a admissão de que existem zonas de incerteza que precisam ser conhecidas.

Num plano metafórico, a pergunta é o indício de que a avaliação se curvará à realidade e de que se parte, de fato, em uma jornada de descoberta crucial para a sobrevivência do objeto, para seu desenvolvimento ou melhoria. Enquanto apostas, projeções, expectativas e riscos são o espírito do planejamento, incerteza, interrogação, crítica e juízo de valor são o espírito da avaliação, o que torna tais práticas componentes gemelares e vertebrais da gestão genuinamente comprometida.

Para Silva (2012b), as perguntas avaliativas devem se ancorar em dois elementos principais, se o objetivo é levar a descobertas e aprendizagem:<sup>17</sup>

*Primeiro [...] devem ser capazes de traduzir aquilo que os atores querem e precisam compreender melhor, vertendo interesses e desconfianças em perguntas. Segundo, devem ser capazes de apontar para direções que possam agregar valor ao objeto avaliado, seja por iluminar suas limitações, seja por reconhecer seus resultados (SILVA, 2012b:11).*

Nas palavras de Davidson (2012:21), “uma avaliação verdadeiramente relevante depende das perguntas certas”,<sup>18</sup> enquanto muitas avaliações se perdem exatamente por não haver clareza sobre o que é realmente importante descobrir. Segundo a autora:

*Uma boa maneira de assegurar a relevância é criar uma lista de perguntas abrangentes para nortear todo o trabalho. Muitas vezes, parte-se diretamente para a coleta ou mensuração de informações, ou as perguntas são limitadas demais, impedindo a obtenção de dados indicativos. Isso ocorre porque se pensa ‘O que se vai mensurar?’, em vez de se pensar ‘O que realmente é preciso saber?’ (DAVIDSON, 2012: 21).*

Observe a Figura 2.3.

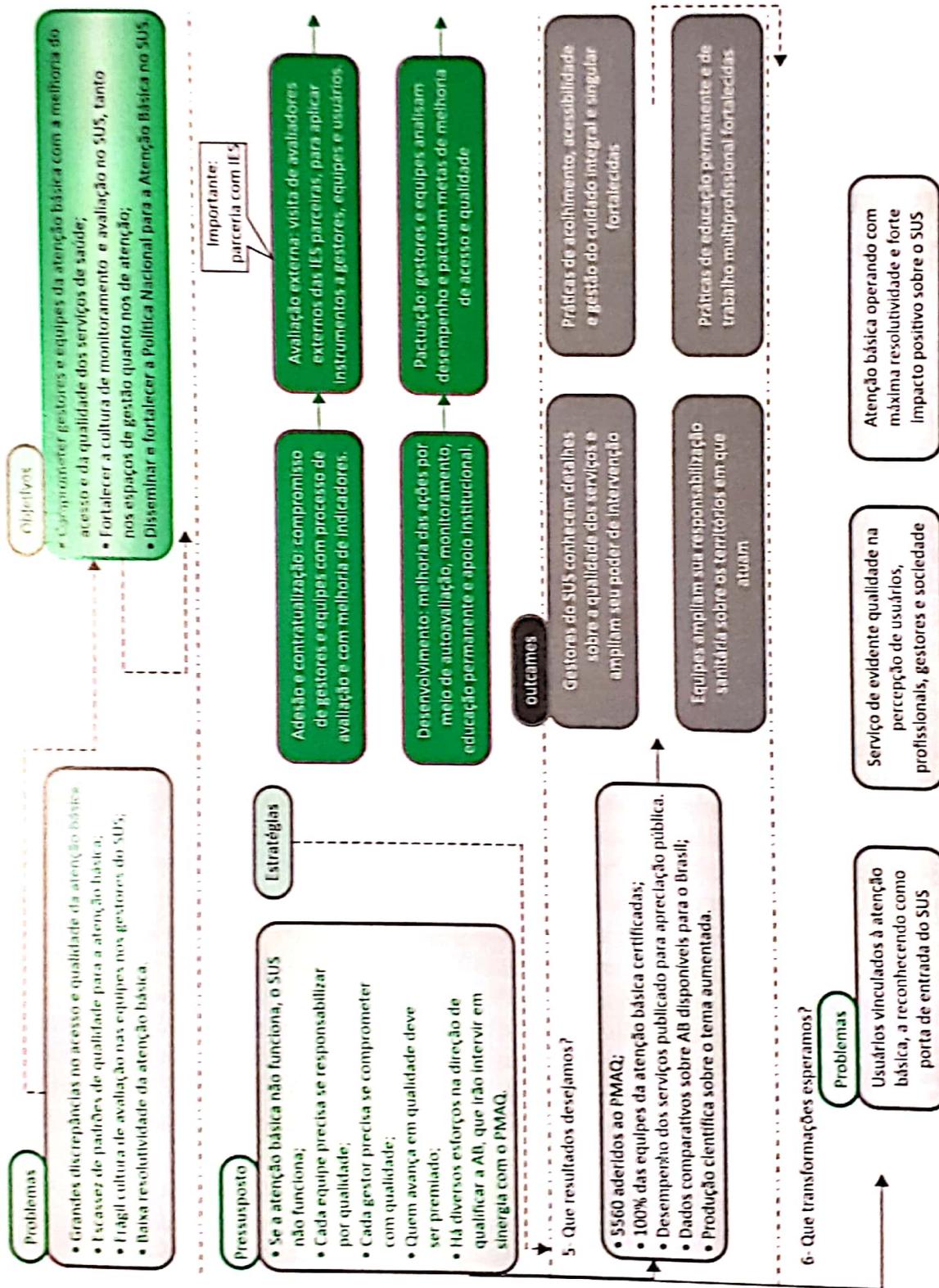


Figura 2.3 – Esquema visual de um programa de saúde. Fonte: SILVA (2015).<sup>16</sup>

Nesse sentido, talvez seja muito importante retomar a conhecida frase de John Tukey, conhecido estatístico norte-americano citado por Davidson (2012:21): “É muito melhor ter uma resposta aproximada para a pergunta certa – que muitas vezes é vaga –, do que uma resposta precisa para a pergunta errada – que pode sempre ser feita de forma precisa”. Como se pode perceber, há uma clara aposta na importância das perguntas como norteadoras de uma avaliação, sendo sua elaboração a etapa primordial de um bom plano de avaliação.

## As matrizes avaliativas como instrumentos de gestão da avaliação

As avaliações são processos que exigem de avaliadores permanentes exercícios lógicos e dialógicos. O exercício lógico se expressa no esforço de conferir coerência técnica ao estudo, de forma que ele alcance os propósitos para os quais foi desenhado. O dialógico se expressa na sustentação de dinâmicas conversacionais significativas, crucial para assegurar que os diferentes atores construam sentido para o processo avaliativo e, assim, o tornem relevante em seus contextos organizacionais.

Como garantir a melhor articulação entre lógica e diálogo numa avaliação? Nossa aposta se concentra em construir a matriz avaliativa com a participação efetiva dos atores. Quando construímos uma matriz, estamos a elaborar o plano de avaliação, ou seja, o caminho técnico capaz de produzir os saberes necessários a respeito de certo objeto. A matriz cumpre o papel de encadear e relacionar os diferentes componentes da avaliação, explicitando sua interdependência para o estudo e permitindo que os atores fiquem atentos a tais conexões.

As matrizes são, também, peças de gestão transparente e de comunicação, uma vez que mostram o que será avaliado, de que forma e que tipo de informação será produzido. O Quadro 2.2 traz um exemplo de matriz avaliativa e apresenta componentes que serão discutidos mais adiante. Cabe destacar que as matrizes avaliativas podem variar em sua forma e componentes, a depender da abordagem teórica e das necessidades do grupo que a constrói. No Quadro 2.2, apresentamos o que consideramos ser conteúdos essenciais.

A matriz tem por finalidade articular a lógica do estudo, sendo capaz de demonstrar os componentes necessários e as etapas a serem cumpridas até a apresentação e debate dos resultados da avaliação, ou seja, até o momento em que o esforço avaliativo se encontra com o momento de decisão e formulação. A matriz também permite visualizar a lógica do estudo, possibilitando que os atores analisem e corrijam sua consistência técnica e reflitam sobre sua viabilidade: quais indicadores melhor respondem à pergunta? Quais fontes de informação são necessárias? Quais formas de coleta de informação são adequadas? Cada etapa subsequente está relacionada à anterior.

Tendo em vista que já abordamos criteriosamente a formulação de perguntas avaliativas e que este capítulo não pretende abordar aspectos relacionados a fontes de informação e formas de coleta, concentraremos o olhar no tema dos indicadores e critérios de julgamento, temas que podem ser considerados o “centro nervoso” das práticas avaliativas. Mas, afinal, o que são indicadores? Por que tanto debate em torno de sua construção? Por que precisamos de critérios de julgamento em uma avaliação? Como construir critérios justos e assegurar uma avaliação justa?

Quadro 2.2 – Exemplo de matriz avaliativa

| Perguntas avaliativas   | Indicadores e critérios                          | Fontes de informação            | Formas de coleta de dados | Produtos da avaliação  |
|---|--|---------------------------------|---------------------------|--|
| O programa de combate à obesidade infantil tem sido implementado de forma sensível ao contexto dos povos indígenas aos quais se dirige? | Linguagem dos materiais utilizados               | Materiais técnicos distribuídos | Leitura e análise técnica | Painel de resultados apresentado para debate no Conselho Estadual de Saúde |
|   | Preparo técnico dos profissionais                | Profissionais de saúde          | Survey                    |  |
|   | Articulação com líderes das aldeias              | Líderes das aldeias alcançadas  | Entrevistas individuais   |  |
|   | Respeito à cultura alimentar dos povos indígenas | População das aldeias           | Grupos focais             |  |

Todo indicador deve ser um resumo da realidade, deve revelar o funcionamento, a condição ou o desempenho de um objeto, deve conter na parte que ele evidencia algo do todo ao que ele se relaciona. Um indicador é, portanto, o fragmento de uma realidade ampla e complexa que não pode ser apreendida ou representada com simplicidade. Por isso, um indicador é sempre um elemento ousado, uma tentativa de colocar uma lupa nos pontos críticos e mais relevantes da realidade, o que exige escolhas antecedidas de bastante reflexão. Tomemos alguns exemplos de indicadores para melhor enxergar tais elementos e refletir sobre eles.

O coeficiente de mortalidade infantil, ou taxa de mortalidade infantil, é um indicador que tem como fórmula de cálculo o número de nascidos vivos que morrem antes de completar 1 ano de vida dividido pelo número total de nascidos vivos. O resultado é, então, multiplicado por mil para se alcançar uma taxa como 14/1000, que é próxima ao último valor divulgado pelo Ministério da Saúde para o Brasil. Esse indicador foi escolhido porque é capaz de apontar algo relevante a respeito das condições de pré-natal, parto e puerpério. Isoladamente, o indicador não revela em qual ou quais dessas etapas há problemas, mas oferece a possibilidade de afirmarmos que as condições de nascimento das crianças em um determinado território são mais ou menos ruins. A taxa de mortalidade infantil é específica e precisa porque mede apenas uma variável. Ela permite comparar diferentes realidades nas quais apliquemos a mesma fórmula, permite comparações ao longo do tempo dentro da mesma população ou entre populações, além de ser uma variável de fácil produção e de bastante tangibilidade. Ao revelar o fato de que crianças que deveriam estar vivas e se desenvolvendo como seres humanos estão mortas antes de completar seu primeiro ano de vida, ele se torna um indicador de forte apelo social e político.

Mas voltemos à Figura 2.1 e tomemos um indicador bem menos tradicional do que o coeficiente de mortalidade infantil: o de respeito à cultura alimentar dos povos indígenas.

Nesse caso, admitimos que um indicador não precisa ser sempre quantitativo. Esse equívoco ocorre em virtude de a própria palavra “indicador” ser frequentemente utilizada em situações qualitativas e mesmo corriqueiras. A face ruborizada de alguém não pode ser um indicador de vergonha? Um sorriso largo não pode indicar alegria? Alguém deixar uma reunião de maneira abrupta não pode indicar descontentamento? Se os indicadores são sinais da realidade, podemos assumir que há sinais de naturezas quantitativa e sinais de natureza qualitativa.

No caso do respeito à cultura alimentar dos povos indígenas, a forma de construir o indicador é outra. Enquanto o indicador de natureza quantitativa demanda uma fórmula e um cálculo, o indicador de natureza qualitativa demanda um texto ou um anúncio que o transforma numa categoria de análise que poderá ser observada na realidade. Pensemos, por exemplo, que o respeito à cultura alimentar dos povos indígenas seja definido da seguinte forma: programa identifica os hábitos alimentares tradicionais dos povos indígenas e incentiva que as discussões sobre alimentação sejam feitas comparando hábitos tradicionais e hábitos recentes adotados a partir do contato com o homem branco, de forma a favorecer padrões de consumo mais equilibrados, sustentáveis, viáveis e que gerem menos obesidade.

Nesse caso, estamos trocando uma fórmula de cálculo (percentual de indígenas obesos ou percentual de indígenas que consomem alimentos tradicionais) pela leitura de como um programa se molda à cultura local no intuito de respeitá-la e valorizá-la. A intenção não é avaliar quais indicadores, se os quantitativos ou os qualitativos, são melhores, julgamento não apenas pobre como estúpido na contemporaneidade. Estamos sustentando que certos objetos, perguntas e circunstâncias necessitam de olhares qualitativos, enquanto outros objetos, perguntas e circunstâncias necessitam de olhares quantitativos. Visto que cada abordagem ilumina certos elementos e enfatiza aspectos distintos, esperamos que seu uso combinado e inteligente eleve a qualidade de uma avaliação.

Sejam quantitativos, sejam qualitativos, os indicadores não são suficientes para que uma avaliação cumpra sua finalidade primária, que é subsidiar a formação de juízo de valor sobre determinado objeto. Para que haja as condições necessárias à elaboração desse juízo de valor e para que esse exercício seja justo e potente, é crucial que avancemos dos indicadores para o que chamamos de “critérios de juízo” ou “critérios de julgamento”. Tomemos como exemplo um programa de combate à doença hipertensiva crônica em uma comunidade quilombola, na qual essa enfermidade está fortemente presente e com dramáticas consequências.

Para avaliar os resultados desse programa, um grupo de trabalho constituído pelo Conselho Municipal de Saúde formulou algumas perguntas avaliativas e, para uma delas, elegeu três indicadores:

- 1) Taxa de acesso a anti-hipertensivos de uso contínuo;
- 2) Participação dos usuários nos grupos de controle de peso;
- 3) Papel dos agentes de saúde na redução do consumo de sal de cozinha (NaCl) na comunidade.

Tais indicadores visam a distintos elementos relacionados ao combate à doença hipertensiva crônica. Enquanto o acesso a medicamentos visa ao controle dos casos medicamentosos e à redução de complicações, a participação dos usuários nos grupos de controle de peso visa à redução de fatores de risco em usuários específicos. Ao olhar o papel dos agentes de saúde na redução do consumo de sal de cozinha, a avaliação observa os potenciais efeitos do programa no longo prazo.

Entretanto, e se descobrirmos que a taxa de acesso a anti-hipertensivos é de 78%? E se descobrirmos que a participação dos usuários nos grupos é intermitente, com pouco vínculo e

muitas faltas? E se descobrimos que os agentes de saúde não estão convencidos de que o sal de cozinha seja um problema e evitem abordar o tema para não parecerem invasivos e pretensiosos? Tais descobertas são, seguramente, importantes, mas elas só ganharão *status* avaliativo na medida em que aquilo que constatamos possa ser comparado com algum parâmetro, possibilitando que formulemos juízos de valor sobre as variáveis e o programa. Aprofundemos o exemplo.

Tendo em vista que os resultados nunca estão inscritos num pano de fundo em branco, sempre há expectativas, promessas e compromissos, mais ou menos explícitos, que constituem um constructo a partir do qual se pode fazer juízo de valor. No Quadro 2.3, apontamos o que poderiam ser as expectativas para cada um dos três indicadores relacionados à avaliação do programa de combate à doença hipertensiva crônica, a fim de deixar claro que o desempenho de um indicador só terá valor se ele encontrar base para comparação. Essa base pode ser proveniente da literatura científica, da experiência concreta dos atores em outros programas similares, da projeção feita no momento de planejamento ou mesmo de um pacto construído pelos atores no processo avaliativo. Venha de onde vier, o elementar é que os

**Quadro 2.3** – Critérios para analisar os indicadores

| Indicadores   | Desempenho satisfatório   | Desempenho razoável   | Desempenho insatisfatório  |
|---|---|---|--|
| Taxa de acesso a anti-hipertensivos de uso contínuo                                     | 91% ou mais dos usuários com diagnóstico de doença hipertensiva crônica têm acesso aos medicamentos prescritos pelo médico de família                                       | Entre 70% e 90% dos usuários com diagnóstico de doença hipertensiva crônica têm acesso aos medicamentos prescritos pelo médico de família   | Menos de 70% dos usuários com diagnóstico de doença hipertensiva crônica têm acesso aos medicamentos prescritos pelo médico de família                             |
| Participação dos usuários nos grupos de controle de peso                                | 80% ou mais dos usuários com IMC acima de 30 participam dos grupos de controle de peso  | Entre 50 e 79% dos usuários com IMC acima de 30 participam dos grupos de controle de peso   | Menos de 50% dos usuários com IMC acima de 30 participam dos grupos de controle de peso  |
| Papel dos agentes de saúde na redução do consumo de sal de cozinha (NaCl) na comunidade | Agentes de saúde estão convencidos da relação entre consumo elevado de NaCl e doença hipertensiva crônica e atuam de maneira firme para reduzir seu consumo pela comunidade | Agentes de saúde estão parcialmente convencidos da relação entre consumo elevado de NaCl e doença hipertensiva crônica, mas não atuam ou quase não atuam para reduzir seu consumo pela comunidade | Agentes de saúde não estão convencidos da relação entre consumo elevado de NaCl e doença hipertensiva crônica e não atuam para reduzir seu consumo pela comunidade |

parâmetros sejam sólidos e justos: nem uma imagem, objetivo impossível de alcançar, nem um desempenho baixo que poderia esconder fragilidades.

Nota-se que os critérios de classificação para cada um dos indicadores são, nesse caso, trabalhados numa escala que oferece três tipos de variação. Nos dois primeiros indicadores, o critério é evidentemente quantitativo, enquanto no terceiro é qualitativo. Como vai se tornando visível, os critérios criam as condições para que os 78% da taxa de acesso a anti-hipertensivos, a participação intermitente dos usuários nos grupos e o não convencimento dos agentes de saúde sobre as consequências do consumo de sal de cozinha sejam mais bem compreendidos e classificados com base em um parâmetro explícito. É desse exercício, que preenche o indicador e critério, que nasce o que essencialmente deveria ser uma avaliação: a aplicação de critérios explícitos e justos, compreendidos e legitimados pelos atores, para julgar a relevância e o mérito de um determinado objeto.

No entanto, a tarefa não cessa aí. O juízo de valor a respeito de um objeto complexo não deveria ser feito com base em apenas dois ou três indicadores, tampouco apartado do contexto no qual o programa se desenvolve; contexto que é político, econômico, cultural etc. Um juízo sólido é capaz de articular os critérios associados a cada indicador, combinando-os da forma mais contextualizada e justa possível. Como dissemos, a avaliação em questão buscava responder outras perguntas e, nesse sentido, o juízo de valor deveria ser composto dos elementos relacionados a todas elas. O Quadro 2.4 mostra a combinação de critérios que possibilitaria um juízo mais cuidadoso. A associação de critérios tem sido denominada rubricas avaliativas e utilizada por diferentes correntes profissionais e autores.<sup>18,19</sup>

O quadro demonstra que o juízo de valor abrange distintos componentes do programa de maneira articulada, convidando os atores a formarem seu juízo de valor de forma contextualizada. Em algumas situações, a prática investigada pode até não corresponder exatamente à especificação de cada critério presente em um mesmo nível da rubrica. Nesse sentido, a rubrica deve ser tomada como a descrição de um domínio de resultado que abrange todos os critérios, sendo necessário o exercício de encontrar o nível de rubrica que melhor representa a situação investigada, permitindo um juízo de valor consciente e sensível à realidade. Uma terceira pergunta poderia, ainda, focalizar o componente econômico do programa, enquanto uma quarta pergunta poderia investigar a forma como a Secretaria de Saúde tem criado efetivas condições para o trabalho das equipes de saúde da família. Lembrando que o contexto político local poderia ser de turbulência, que poderia estar em curso uma forte crise de arrecadação, que profissionais poderiam ter sido aprovados em um concurso público no município vizinho e deixado o trabalho, entre outros, o juízo de valor deveria abranger esse conjunto para formular afirmações precisas sobre o programa e favorecer que decisões efetivamente estratégicas fossem tomadas com base na avaliação. O exercício de avaliar é, nesse sentido, um exercício de precisão e de justiça, de sensibilidade e abrangência, de transparência e compromisso, características essenciais para a gestão que desejamos para os serviços de saúde.

## Considerações finais

Não restam dúvidas de que a gestão é um campo de enormes desafios. A ela cabe a inteligente e extenuante tarefa de equilibrar interesses a fim de garantir avanços e manejar dinâmicas de poder para favorecer que os serviços cumpram seus papéis sociais de proteger, recuperar e promover a saúde da população. Sustentar os ciclos de planejamento, implementação,

Quadro 2.4 – Rubricas avaliativas para permitir juízo de valor.

| Perguntas  | Programa tem desempenho satisfatório e deve ser continuado para garantir impactos   | Programa tem desempenho razoável e precisa de correções pontuais para melhorar entregas   | Programa tem desempenho insatisfatório e precisa ser redesenhado   |
|--|---|---|--|
| As estratégias do programa têm incidido de maneira potente nos principais fatores de risco para a população quilombola coberta?  | 91% ou mais dos usuários com diagnóstico de doença hipertensiva crônica têm acesso a medicamentos; 80% ou mais dos usuários com IMC acima de 30 participam dos grupos de controle de peso; e os agentes de saúde estão convencidos da relação entre consumo elevado de NaCl e doença hipertensiva crônica e atuam de maneira firme para reduzir seu consumo pela comunidade | Entre 70 e 80% dos usuários com diagnóstico de doença hipertensiva crônica têm acesso a medicamentos; entre 50 e 70% dos usuários com IMC acima de 30 participam dos grupos de controle de peso; e os agentes de saúde estão parcialmente convencidos da relação entre consumo elevado de NaCl e doença hipertensiva crônica, mas não atuam ou quase não atuam para reduzir seu consumo pela comunidade | Menos de 70% dos usuários com diagnóstico de doença hipertensiva crônica têm acesso a medicamentos; menos de 50% dos usuários com IMC acima de 30 participam dos grupos de controle de peso; e os agentes de saúde não estão convencidos da relação entre consumo elevado de NaCl e doença hipertensiva crônica e não atuam para reduzir seu consumo pela comunidade |
| As equipes de saúde da família têm sido capazes de compreender e adaptar os protocolos de cuidado de forma a ampliar sua capacidade de detecção precoce e manejo dos casos de doença hipertensiva? | As equipes conhecem, compreendem e utilizam os protocolos de maneira adaptada ao contexto local e têm sido capazes de ampliar a detecção precoce de doença hipertensiva e de atuar para evitar complicações   | As equipes conhecem, compreendem os protocolos, mas o utilizam de forma parcial e têm sido pouco capazes de ampliar a detecção precoce de doença hipertensiva e de atuar para evitar complicações   | As equipes conhecem, mas pouco compreendem ou se importam com os protocolos, sendo incapazes de ampliar a detecção precoce de doença hipertensiva e de atuar para evitar complicações  |

monitoramento e avaliação é uma ação de extrema relevância para ampliar acesso e qualidade dos serviços, o que requer a utilização de nossas subjetividades, nos pede inspiração, exige transparência, demanda lógica e organização e requer investimento permanente em boas conversas.

No entanto, é sempre oportuno refletir que os papéis técnicos e políticos atribuídos aos gestores pode degenerar-se em um centralismo exagerado, verticalizado e autoritário, prejudicando a criação de espaços em que os diversos atores sejam convidados a participar e onde as diferentes perspectivas, experiências e conhecimentos sejam valorizados. Acreditamos que a criação desses espaços deva ser incentivada no trabalho em saúde, proporcionando um horizonte em que a cogestão, no sentido atribuído por Campos,<sup>20</sup> não apenas subsidiando processos avaliativos mais democráticos e responsivos, mas também contribua para que esses espaços de trabalho sejam dotados de sentido para todos que neles atuam. Ainda que tal horizonte, muitas vezes, pareça distante, já que as intensas exigências e desafios que o cotidiano nos impõe acabam minando nossa potência criativa, o trabalho em saúde é sempre um fazer em saúde, cujas interações humanas constituem motor de transformações que alargam as possibilidades à medida que são colocadas em prática, mesmo que sob constante tensão.

Nesse movimento complexo, as práticas de avaliação podem operar como potente dispositivo de implicação, descoberta, aprendizagem e apoio à mudança. Como pudemos observar na argumentação deste ensaio e também ao longo de toda a obra, somos parte de uma comunidade de sujeitos que acredita no olhar criterioso e reflexivo sobre a trajetória das organizações, sobre suas apostas e conquistas, falhas e limites. Tomamos as avaliações ora como mesas de negociação, ora como assembleias de concertação. Lançamo-nos nos processos ora como facilitadores de conversas, ora como investigadores interessados nos detalhes. E está nessa plasticidade a potência das avaliações.

Entretanto, não carregamos ilusões ingênuas. Sabemos das cargas históricas despejadas sobre as avaliações desde as remotas e ainda resilientes construções de comando, controle, padronização e massificação legadas por Fayol e Taylor. Sabemos o quanto as avaliações foram utilizadas ao longo do século XX como potentes exercícios de poder e de cerceamento de possibilidades. Sabemos quantas vezes os conceitos ciência, neutralidade e objetividade foram utilizados para conservar o poder nas mãos de poucos e submeter as multidões. Sabemos quanta violência simbólica ainda reveste os processos avaliativos e sabemos que as tentativas de produzir avaliações justas numa moldura democrática requerem forte disposição para disputar o espaço e convencer pela experiência.

Esperamos que este capítulo seja tomado como mais uma contribuição para promover e fortalecer o direito à saúde, que só pode ser garantido de maneira adequada mediante o desenvolvimento de serviços de saúde desenhados e executados para terem relevância e produzirem efeitos positivos na sociedade. Não está entre nossas premissas a que considera as práticas avaliativas a panaceia da gestão ou tábua de salvação das políticas públicas; o mundo é bem mais complexo do que isso.

Contudo, vive em nós a premissa de que não haverá universalidade, equidade e integralidade no SUS se os serviços e os programas não avançarem em sua relevância, ou seja, se não forem concebidos e desenhados com grande capacidade de aderirem à realidade, de se abraçarem aos profissionais de saúde, de lançarem mão das tecnologias mais inteligentes e efetivas e de se sustentarem financeiramente. Não haverá universalidade, equidade e integralidade no SUS se os serviços e programas não tiverem êxito em seus propósitos de alterar positivamente as condições de saúde na sociedade. O que desejamos é que as práticas avaliativas apoiem e encorajem os diversos atores que fazem o SUS a conhecer, analisar, aprender e ajustar permanentemente a relevância e o mérito dos serviços.

## Referências Bibliográficas

1. Guba FG, Lincoln YS. *Avaliação de Quarta Geração*. Campinas: Editora da Unicamp, 2011.
2. House ER. The role of the evaluator in a political world. *The Canadian Journal of Program Evaluation*, v. 19, n. 2, p. 1-16, 2004.
3. Cousins JB, Whitmore E. Framing Participatory Evaluation. In: Whitmore, E (ed.) *Understanding and Practicing Participatory Evaluation. New Directions for Evaluation*, no. 80. San Francisco, 1998.
4. Mendes MFM, Bezerra GCLCA, Dubeux LS. Avaliabilidade ou pré-avaliação de um programa. In: SAMICO, I. et al. (org.). *Avaliação em saúde – bases conceituais e operacionais*. Recife: IMIP, 2010.
5. Wholey JS. Evaluability assessment: developing program theory. *New Directions for Program Evaluation*, n. 33, p. 77-92, 1987.
6. Vieira da Silva LM. *Avaliação de políticas e programas de saúde*. Rio de Janeiro: Fiocruz, 2014.
7. Medina FG et al. Uso de modelos teóricos na avaliação em saúde: aspectos conceituais e operacionais. In: Hartz ZMA, Silva LM Vieira da. (Org.). *Avaliação em saúde: dos modelos teóricos à prática na avaliação de programas e sistemas de saúde*. Salvador: EDUFBA, 2005.
8. Medeiros PFP, Bezerra LCA, Santos NTV, de Oliveira Melo E. Um estudo sobre a avaliabilidade do Programa + Vida: política de redução de danos em álcool, fumo e outras drogas do município de Recife, Brasil. *Revista Brasileira de Saúde Materno Infantil*, v. 10, supl. 1, p. s209-s217, nov. 2010.
9. Padilha MA, Oliveira CM, Figueiró AC. Estudo de avaliabilidade do Programa Academia Carioca da Saúde: desafios para a promoção da saúde. *Saúde em Debate*, v. 39, n. 105, p. 375-386, jun. 2015.
10. Oliveira CM, Cruz MM, Kanso S, Reis AC, Lima A, Torres RMC, et al. Avaliabilidade do Programa de Valorização do Profissional da Atenção Básica (PROVAB): desafios para gestão do trabalho. *Ciência e Saúde Coletiva*, v. 20, n. 10, p. 2999-3010, out. 2015.
11. Cassiolato M, Guerresi S. *Como elaborar modelo lógico: roteiro para formular programas e organizar avaliação*. Brasília, 2010.
12. BRASIL. Ministério do Desenvolvimento Social e Combate à Fome. *Caderno de Estudos: curso em conceitos e instrumentos para o monitoramento de programas*. Brasília, 2014. 91.
13. Chen TH. *Theory-Driven Evaluations*. Beverly Hills: Sage, 1990.
14. Cruz MM. *Avaliação de Programas de Prevenção de DST/AIDS para Jovens: estudo de caso numa organização governamental e numa organização não-governamental do Município do Rio de Janeiro*. 2006. 234 f. Tese (Doutorado em Saúde Pública) – Escola Nacional de Saúde Pública Sérgio Arouca (ENSP), Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, 2006.
15. Silva RR. Sobre avaliação e desenvolvimento organizacional. In: OTERO, M. (Org.). *Contexto e prática da avaliação de iniciativas sociais no Brasil: temas atuais*. São Paulo: Instituto Fonte/Peirópolis, 2012a. p. 43-60.
16. Silva RR, Furtado J, Akermab M, Gasparini M. Subsídios para a meta-avaliação do Programa de Melhoria do Acesso e Qualidade do SUS PMAQ. In: Furtado J, Akerman M. *Práticas de avaliação em saúde no Brasil: diálogos*. Porto Alegre: Editora Rede Unida, 2015.
17. Silva RR. Para fazer avaliações relevantes. In: Fundação Itaú Social, Fundação Roberto Marinho, Move (orgs.) *A Relevância da avaliação para o investimento social privado*. São Paulo: Fundação Santillana, 2012b.p. 09-16.
18. Davidson EJ. Tornar as avaliações estrategicamente práticas e relevantes. In: Fundação Itaú Social, Fundação Roberto Marinho, Move (orgs.) *A Relevância da avaliação para o investimento social privado*. São Paulo: Fundação Santillana, 2012.p. 17-41.
19. Chianca TK. Avaliações válidas, relevantes e úteis. In: Fundação Itaú Social, Fundação Roberto Marinho, Fundação Maria Cecília Souto Vidigal, Move (orgs.). *Avaliação para o investimento social privado: metodologias*. São Paulo: Fundação Santillana, 2013. p. 93-116.
20. Campos GWS. Clínica e saúde coletiva compartilhadas: teoria Paidéia e reformulação ampliada do trabalho em saúde. In: Campos GWS, Figueiredo MD, Pereira Júnior N, Castro CP (orgs.). *Tratado de saúde coletiva*. São Paulo; Rio de Janeiro: Hucitec; Fiocruz, 2006. p. 53-92.