

PREFÁCIO

Esperamos que você, leitor, considere este livro sensivelmente diferente de qualquer outro livro sobre avaliação que já tenha lido. Não descrevemos o campo da avaliação do modo como ele surgiu historicamente, mas nos referimos a alguns fatos históricos aqui e ali. Não abordamos os vários diferentes modelos empregados na atividade de avaliação, cujas publicações são abundantes, embora, de vez em quando, citemos brevemente determinados modelos. Não examinamos a avaliação principalmente como um processo técnico de investigação, enfatizando ao mesmo tempo as técnicas, em especial as técnicas estatísticas, postura que se poderia esperar de um livro como este, conquanto sem dúvida abordemos problemas processuais, tanto quanto a proposição de critérios de adequação que nos parecem aplicáveis. Talvez mais surpreendente, não examinamos a avaliação como um processo *científico*, porque estamos convictos de que abordar a avaliação do ponto de vista científico significa perder por completo seu caráter fundamentalmente social, político e centrado em valores. Nossa intenção é definir uma abordagem emergente mas madura da avaliação, que vá além da mera ciência — não se restrinja à obtenção de fatos — para abranger a miríade de elementos humanos, políticos, sociais, culturais e contextuais que ela encerra. Chamamos essa nova abordagem de *avaliação de quarta geração* para assinalar que nossa construção transpõe as gerações previamente existentes, centradas na mensuração, na descrição e no juízo de valor, para abarcar um novo nível em que a principal dinâmica é a *negociação*. Esse novo modelo tem algumas características interessantes.

Primeiro, esse modelo assume o ponto de vista de que os resultados da avaliação não são descrições do “modo como as coisas de fato são” ou “de fato funcionam” ou de algum estado de coisas “verdadeiro”. Na realidade, eles representam construções significativas que atores específicos ou vários grupos de atores formam para “atribuir sentido” às situações nas quais eles se encontram. Os resultados não são “fatos” em um sentido definitivo. Na realidade, eles são *criados* literalmente por meio de um processo interativo que *inclui* o avaliador (chega de objetividade!), bem como os vários grupos de interesse ou interessados (*stakeholders*) que são postos em risco pela avaliação. Desse processo surgem uma ou mais *construções* que são as realidades daquele caso.

Segundo, ele reconhece que as construções por meio das quais as pessoas atribuem sentido às suas situações e condições são em grande medida moldadas pelos valores dos construtores. Evidentemente, se todos os construtores compartilharem um mesmo sistema de valores, haverá poucos problemas. Aliás, aparentemente, pode ser que nessas circunstâncias a construção compartilhada entre todos *seja* o verdadeiro estado de coisas, caso em não há praticamente nenhum prejuízo para aqueles que se envolvem em uma ilusão desse tipo. Contudo, quase todas as sociedades modernas (incluindo aquelas supostamente “primitivas” encontradas, digamos, nos países do terceiro mundo) demonstram ter uma estrutura de valor pluralista. Portanto, a questão sobre os valores que devem ser considerados e como diferentes pontos de vista podem ser acomodados torna-se soberana. É imediatamente evidente e provável que a metodologia que se considera isenta de valores terá pouca utilidade nesses casos.

Terceiro, ele propõe que essas construções estão intrinsecamente vinculadas a contextos físicos, psicológicos, sociais e culturais específicos nos quais elas são formadas e aos quais elas se referem. O contexto apresenta o “entorno” no qual as pessoas que formam as construções vivem e ao qual elas tentam atribuir sentido. Ao mesmo tempo, esse entorno mantém-se informe até o momento em que as construções das pessoas que nele se encontram dotam-no de parâmetros, características e limites. Em um sentido bastante literal, o contexto vivifica e é vivificado pelas construções que as pessoas vêm a formar e a defender. Visto que essas mesmas pessoas formam parte do contexto umas das outras, é compreensível que se desenvolvam construções compartilhadas no decorrer do tempo; aqueles que

pertencem a um determinado contexto chegam a um consenso sobre sua natureza. Entretanto, o fato de haver consenso não significa que exista um grau maior de realidade em relação a qualquer coisa que venha a ser consentida em conjunto; significa apenas que as pessoas que estão de acordo alcançaram uma construção conjunta que é real para elas. Além disso, o consenso é produto de uma troca conceitual humana em um ambiente *particular*; portanto, é improvável que esse mesmo consenso necessariamente possa ajudar outras pessoas a atribuir sentidos a *seu* ambiente. Em suma, é precisamente pelo fato de serem produto do pensamento humano que as construções consensuais estão sujeitas a erros humanos; nenhuma delas pode ser considerada “verdadeira” em sentido absoluto, nem mesmo uma aproximação da “verdade”.

Quarto, essa forma emergente de avaliação reconhece que as avaliações podem ser moldadas de várias maneiras para conceder direitos aos grupos de interesse ou os privar de direitos. É claro que pode haver um envolvimento seletivo desses interessados no projeto e na implementação da avaliação. Se apenas o avaliador e o cliente tiverem o privilégio de decidir sobre as perguntas a serem colocadas, a instrumentação a ser empregada, o método de análise e interpretação de dados a ser usado, e outros, será negada a outros interessados a oportunidade de buscar interesses próprios legítimos. Em vista da “relação amistosa” (somos gratos a Michael Scriven por esse termo) que em geral existe entre o avaliador e o cliente, não é de surpreender que as avaliações tendam a revelar ineficácias e prevaricações em grupos que *não* sejam os clientes (por exemplo, os professores que são responsabilizados pelas supostas condições lamentáveis das escolas) ou que os objetivos do cliente sejam considerados mais prioritários que os dos demais interessados (por exemplo, na avaliação de procedimentos policiais alternativos para intervir em disputas domésticas com base no critério de menor reincidência, uma situação que a polícia almeja porque reduz riscos, trabalhos administrativos, comparecimentos em tribunais, entre outras vantagens, e não no critério de violência contra a mulher, que tem poucos impactos imediatos na carga de trabalho da polícia). Os interessados podem receber poder ou ser privados de poder por meio da disseminação seletiva dos resultados de avaliação. Se informação é poder, então negá-la significa diminuir o poder. Se os clientes tiverem a palavra final em relação às informações que serão divulgadas, a quem, quando e

por quais meios, o processo sem dúvida tenderá a manter e mesmo a ampliar o poder daqueles que já o detêm e, de maneira concomitante, a tirar das pessoas relativamente impotentes até mesmo o pouco poder que elas têm.

Quinto, ele propõe que a avaliação deve centrar-se em uma ação que defina o rumo a ser seguido, estimule os interessados a segui-lo e gere e preserve seu compromisso em proceder dessa forma. Os avaliadores, talvez à imagem de seus confrades mais antigos, os cientistas “puros”, parecem inclinados a relegar a outros o acompanhamento (*follow-up*), o aspecto aplicado da avaliação. Que esse acompanhamento nem sempre ocorra já está bem documentado no atual clamor pela utilização da avaliação. Os profissionais dessa área, chocados com a não utilização da avaliação, ora culpam os clientes pelo capricho obstinado em não agir, não obstante a lógica convincente das recomendações, ora se culpam por não “promover” satisfatoriamente o produto da avaliação. Aliás, em seu discurso de posse na Associação Americana de Avaliação, Michael Quinn Patton exortou seus membros a adotar as posturas e técnicas utilizadas pelos vendedores, advertindo-os, no entanto, de que o primeiro requisito para conseguir realizar uma venda é “ter um produto no qual se possa acreditar”. Na maioria das vezes, o “produto” da avaliação é um conjunto de recomendações que serve apenas aos propósitos do avaliador e do cliente, presta pouca atenção aos interesses legítimos de outros grupos de interesse, é indiferente às questões e aos problemas levantados por outros grupos e reflete somente um (o deles) dos vários conjuntos de valores inerentes à situação. Se houver a intenção de ter uma linha de ação com a qual a maioria dos interessados concorde, esse procedimento só será possível por meio da *negociação* que respeite os diferentes conjuntos de valores e possibilite que os indivíduos encontrem um motivo para apoiá-lo, para se dedicar a ele e para se sentirem bem com isso. Se houver intenção de atingir esse objetivo, o avaliador deverá desempenhar uma função bem mais abrangente do que simplesmente a função de técnico que coleta informações; em vigor, ele deve ser o orquestrador do processo de negociação, e em última análise é justamente essa a essência da avaliação.

Em conclusão, a avaliação de quarta geração sustenta que, na medida em que a avaliação envolva seres humanos (como clientes, grupos de in-

teresse, fontes de informação e inúmeras outras entidades), é responsabilidade do avaliador interagir com esses seres humanos de uma maneira que respeite sua dignidade, sua integridade e sua privacidade. É evidente que os avaliadores convencionais têm sido criteriosos com relação à ética; os cânones referentes ao consentimento totalmente informado, aos danos, ao engano e à privacidade/confidencialidade são bem compreendidos e na maioria das vezes respeitados. Entretanto, o pleno significado da frase “respeitar sua dignidade, sua integridade e sua privacidade” não se restringe nem um pouco a esses amparos convencionais. Esse significado eleva-se ao nível do envolvimento *totalmente participativo*, em que os grupos de interesse e outras pessoas que venham a participar da avaliação são acolhidos como parceiros equitativos em todos os aspectos do projeto, da implementação, da interpretação e da ação que resulta da avaliação — isto é, eles recebem uma medida exata de *paridade e controle político*. Isso significa também que os participantes humanos têm o privilégio de compartilhar suas construções e de trabalhar em direção a uma construção *conjunta* comum, consensual e mais fundamentada e esclarecida — eles recebem uma medida exata de *paridade conceitual*. E, evidentemente, isso significa que os participantes continuam a ser tratados como seres humanos, e *não* como sujeitos da experimentação ou objetos de estudo.

Para que as intenções e o compromisso dessa concepção emergente de avaliação sejam cumpridos, para conduzi-la, é necessário encontrar uma forma que reconheça que seus resultados são construídos, uma forma que leve em conta diferentes valores e diferentes contextos (físicos, psicológicos, sociais e culturais), outorgue poder e conceda direitos, amalgame o ato da avaliação e respectivas atividades de acompanhamento em um todo indistinguível e seja totalmente participativa porque estende a paridade política e conceitual a todos os grupos de interesse. Acreditamos que o modelo de avaliação de quarta geração por nós proposto atenda a esse desafio, pelo menos em um primeiro nível de aproximação. É claro que um maior avanço nesse sentido é em grande medida essencial.

A avaliação de quarta geração, como mostraremos, está fundamentada em dois elementos: em um *enfoque responsivo* — determinar quais perguntas devem ser colocadas e quais informações devem ser coletadas com base nas informações fornecidas pelos grupos de interesse — e em uma

metodologia construtivista — conduzir o processo de investigação de acordo com pressuposições do paradigma construtivista*.

Ao longo da história, as avaliações adotaram vários enfoques: encarregar-se da compilação de notas e normas de testes aplicados, normalmente padronizados; avaliar a congruência do desempenho com objetivos pre-determinados; analisar as decisões a serem tomadas e coletar as informações necessárias para fundamentá-las; determinar a utilidade de custo-benefício do sujeito-objeto da avaliação (*evaluand*); avaliar os efeitos de uma inovação ou intervenção, seja ela intencional ou não; determinar o grau com que as necessidades não atendidas anteriormente são satisfeitas ou atenuadas; e outros. Todas essas abordagens se baseiam no compromisso de obter intensa colaboração entre avaliador e cliente (o qual especifica ou, no mínimo, aprova os testes a serem aplicados, os objetivos a serem atingidos, as decisões a serem apoiadas, a intervenção ou inovação a ser testada, as necessidades a serem atendidas e assim por diante), os quais tendem a ignorar outros atores na situação e, particularmente, seus valores e suas construções. A avaliação responsiva foi assim denominada por seu autor, Robert Stake, para enfatizar a ideia de que todos os grupos de interesse colocados em risco por uma avaliação têm o direito de apresentar suas reivindicações, preocupações e questões para que sejam consideradas (respondidas), independentemente do sistema de valores que eles adotam. Ela foi criada como antítese da avaliação preordenada ou predeterminada, que pressupõe que o avaliador e o cliente tenham, em conjunto, informações e legitimação para estruturar e implementar completamente uma

* Em nossos escritos anteriores, com frequência utilizamos os termos *naturalístico* e *naturalista* em referência ao processo de investigação que adotamos e à pessoa envolvida nesse processo. Por uma série de motivos — por exemplo, esses termos evocam o naturalismo britânico do século XIX com o qual não temos nada em comum; eles propõem uma ontologia realista que particularmente rejeitamos; eles implicam uma alegação de legitimidade em relação ao nosso trabalho (“ao modo da natureza”) que é injustificável; e eles levam a crer que estamos adotando uma investigação em ambientes naturais como *solução* para os dilemas paradigmáticos —, sentimos que esses termos não são mais apropriados. Além disso, conseguimos reconhecer que a principal característica de nosso paradigma é a suposição de que as realidades, certamente as realidades sociais/comportamentais, são construções mentais. Portanto, optamos por empregar os termos *construtivismo* e *construtivista* em referência ao paradigma e à pessoa envolvida com sua aplicação, respectivamente. Essa convenção é seguida na maior parte deste livro, exceto nas circunstâncias em que nos referimos ao nosso livro anterior, algumas vezes em citações extraídas desse trabalho. Nesses casos, procuramos lembrar os leitores dessas novas convenções.

avaliação, sem precisar consultar outras partes (embora, na prática, essa consulta com frequência ocorra por motivos políticos).

O segundo elemento de apoio da avaliação de quarta geração é a metodologia construtivista. No passado, a metodologia empregada nas avaliações era quase exclusivamente científica, fundamentada *ontologicamente* na suposição positivista de que existe uma realidade objetiva governada por leis naturais imutáveis e *epistemologicamente* na suposição congênere de que existe uma dualidade entre o observador e o observado que possibilita que o observador se posicione *fora* da esfera do observado, não influenciando nem sendo influenciado por ele. Em vista dessas suposições, o avaliador pode alegar que produziu resultados que representam “as coisas como elas de fato são” e “as coisas como elas de fato funcionam”, não havendo nenhuma possibilidade de influência por parte do avaliador, do cliente ou de outra pessoa qualquer.

A postura da avaliação de quarta geração é bem diferente. Ela parte da suposição de que as realidades não estão objetivamente “lá fora”, mas são construídas pelas pessoas, normalmente sob a influência de uma série de fatores sociais e culturais que geram construções compartilhadas. Entretanto, as construções socialmente compartilhadas não equivalem à “realidade” do positivista; *não existe realidade*, exceto aquela criada pelas pessoas quando elas tentam “atribuir sentido” ao seu entorno. É evidente que essas realidades socialmente construídas, além de *não* serem independentes do “observador” (leitor, “construtor”), dependem totalmente dele para toda e qualquer existência que possam ter.

A metodologia de avaliação que reflete essas suposições deve estar fundamentada em um paradigma de investigação radicalmente distinto do positivismo científico. Acreditamos que esse paradigma exista naquilo que passamos a chamar de paradigma construtivista (também chamado, com diferentes nuanças de significado, de paradigma *interpretativo* ou *hermenêutico* e, algumas vezes — equivocadamente, julgamos —, de paradigma *qualitativo*). Esse paradigma apoia-se em uma ontologia relativista, e não realista, e em uma epistemologia monista e subjetiva, e não dualista e objetiva. Seu exercício une o avaliador e os interessados em uma interação que *cria* o produto da avaliação, utilizando uma abordagem hermenêutico-dialética cujo objetivo é estabelecer essa interação e mantê-la dentro de determinados limites de qualidade. Além disso, o produto da

avaliação não é, em acentuado contraste com a metodologia convencional, um conjunto de conclusões, recomendações ou julgamentos de valor, mas, na realidade, uma *agenda de negociação* das reivindicações, preocupações e questões que não foram resolvidas nas trocas hermenêutico-dialéticas.

Essas alegações, introduzidas neste prefácio da forma mais parcimoniosa possível, são substanciais e não contam com nenhum apoio. Sua explicação e justificação, obviamente, são o objetivo deste livro. Ao redigi-lo, tínhamos em mente vários públicos distintos — e também alguns públicos que exortaríamos a *não* o ler.

Em primeiro lugar, dirigimo-nos aos avaliadores enquanto grupo profissional. Que a avaliação é uma profissão, só se tornou evidente nas últimas décadas, se é que não tenha sido evidente antes. Primeiro, foram criadas duas associações profissionais independentes, a Evaluation Network e a Sociedade de Pesquisa sobre Avaliação (Evaluation Research Society — ERS), que nos últimos três anos se fundiram em um único grupo, a Associação Americana de Avaliação. Além disso, foram criados vários padrões para orientar o trabalho dos avaliadores (e, ao mesmo tempo, informar clientes e usuários). Os mais proeminentes são os Standards for the evaluation of educational programs, projects, and materials [Padrões de avaliação de programas, projetos e conteúdos educacionais] (1981), do Comitê Misto, e seu congênere Personnel evaluation standards [Padrões de avaliação de pessoal] (1988). Por fim, criaram-se inúmeros periódicos, como o *Evaluation Practice*, *Evaluation Review*, *Educational Evaluation and Policy Analysis* e a série *New directions in program evaluation*, por meio dos quais se tem desenvolvido rapidamente uma literatura significativa. Tal como nossa experiência demonstra, é cada vez maior o número de profissionais insatisfeitos com os métodos de avaliação convencionais. Essas pessoas poderão encontrar neste livro um fundamento teórico-filosófico, enquanto alternativa, e sugestões práticas para conduzir uma avaliação. Embora os pesquisadores e analistas de políticas talvez não considerem a avaliação essencial a seus interesses, eles podem encontrar neste livro conteúdos e conceitos úteis para reconstruir os paradigmas de investigação que eles utilizam como orientação.

Este livro destina-se igualmente a grupos de interesse, aquelas pessoas que, quer queiram, quer não, algumas vezes são atraídas para uma avaliação porque têm algum interesse em jogo e, por isso, são colocadas em risco. É

inquestionável, deve-se concluir, que sempre haverá interessados cujas reivindicações, preocupações e questões mereçam ser identificadas e respeitadas. Estão incluídos nesse grupo os *representantes* ou *agentes* que idealizam, atuam, gerenciam, financiam, supervisionam ou, de outra forma, contribuem para o desenvolvimento, a instituição e a utilização do sujeito-objeto da avaliação; os supostos *beneficiários* (normalmente chamados de “públicos-alvo”), que tiram proveito ou devem tirar proveito do sujeito-objeto da avaliação, e os beneficiários secundários que tenham alguma relação com eles — por exemplo, cônjuges de pessoas que estão se submetendo a tratamentos de reabilitação contra alcoolismo, pais de crianças que estão recebendo reforço na escola, entre outros casos; e as *vítimas*, pessoas prejudicadas ou privadas direta ou indiretamente de algum bem pela implementação do objeto de avaliação, e nisso se incluem as oportunidades que foram renunciadas. Os grupos de interesse normalmente não percebem que uma avaliação que os afeta está em andamento ou que estão sendo omitidas informações talvez importantes para eles ou ainda que apenas as perguntas relacionadas a outros grupos, eventualmente concorrentes, estão sendo levantadas. Em geral, os grupos de interesse não têm ideia do papel que podem desempenhar em uma determinada avaliação ou mesmo da legitimidade de seu direito de participação. Esperamos oferecer a esses grupos de interesse uma clara percepção das funções que eles podem desempenhar e das perguntas que eles podem levantar, bem como evidenciar que eles têm direito de se tornar parceiros plenos na atividade de avaliação, tanto política quanto conceitualmente.

Esperamos também atrair clientes ou patrocinadores das atividades de avaliação, isto é, aquelas pessoas que encomendam as avaliações e/ou oferecem recursos financeiros e outros tipos de recurso para a sua realização. A maioria dos clientes ou dos patrocinadores não tem uma ideia muito clara sobre o que devem esperar de uma avaliação; temos esperança de que este livro oferece alguns esclarecimentos. Todavia, gostaríamos principalmente de convencer esses clientes ou patrocinadores de que o modelo convencional de avaliação, que em vigor lhes reserva poder e autoridade na tomada de decisões, não é apenas moral e eticamente injusto e equivocado; ele é também politicamente ingênuo e conceitualmente limitado. As avaliações convencionais raras vezes amalgamam as opiniões públicas por trás da linha de ação que os clientes e patrocinadores consi-

deram sensata; aliás, muito ao contrário — as avaliações tendem a persuadir os interessados de que o principal propósito da avaliação é coletar informações que racionalizam e legitimam alguma decisão já tomada “nos bastidores”. A contribuição pública em geral é solicitada, mas raramente respeitada. Se este livro conseguir revelar alguns dos motivos que justificam a incapacidade da avaliação convencional não apenas de produzir informações úteis, mas também de obter consenso político, teremos conseguido atingir nossos objetivos.

Embora seja nossa expectativa influenciar esses três públicos amplos com todas as coisas que temos a dizer, algumas pessoas, acreditamos, não tirarão proveito da leitura deste livro. Nós a retratamos como pessoas “sectárias”, pessoas convencidas de que já possuem o método para administrar uma avaliação. Michael Quinn Patton fala de uma experiência que ele teve ao conduzir uma avaliação em um vilarejo caribenho. Sua investigação o levou a passar um dia em uma lavoura de cana-de-açúcar na época mais quente do verão, onde um grupo de aldeões estava cortando cana — com certeza um dos trabalhos mais fisicamente exaustivos que se possa imaginar. Para aliviar o calor, os aldeões, homens e também mulheres, estavam praticamente nus. Naquele momento vestiam apenas uma espécie de tanga. De repente, como se tivessem recebido um sinal, as mulheres jogaram o machete ao chão e correram para o outro lado da lavoura em que haviam deixado seus pertences, vestindo rapidamente uma camiseta, um vestido caseiro ou o que tivessem à mão para cobrir os seios nus. Patter só se deu conta disso quando percebeu o que motivava esse estranho comportamento. O padre da aldeia estava vindo de jipe pela estrada. Os aldeões, de orelha em pé, já o tinham ouvido e visto bem antes de Patton tomar consciência disso. Sabendo que o padre consideraria a nudez um motivo suficiente para no mínimo um sermão e talvez para impor uma penitência, as mulheres decidiram sensatamente cobrir os seios. Contudo, tão logo o padre se foi, elas novamente se despiram e retornaram ao trabalho.

Para Patton, pelo fato de os cortadores de cana terem adotado não apenas o catolicismo, mas também a prática de vodu e possivelmente outras crenças, eles não se incomodavam em se vestir — ou em se despir — de acordo com o que fosse apropriado para as circunstâncias. O padre, entretanto, um sectário, não podia aceitar a ideia de que práticas não toleradas

por seu sistema de crenças pudessem ser outra coisa senão um comportamento ruim que merecia condenação e punição.

Para nós, isso se aplica aos partidários fiéis do positivismo. Convictos de que existe uma realidade única e verdadeira, governada por leis naturais, prontas para serem reveladas e utilizadas pelos métodos da ciência, os positivistas rejeitam todos os pontos de vista relativistas, e o construtivismo é um deles, considerando-os não apenas gravemente errados, mas perniciosos e repugnantes. Os defensores desse ponto de vista, pensam eles, por pouco não se enquadram entre os charlatões e vendedores de óleo de cobra. Os sectários provavelmente não encontrarão nada neste livro que possa mudar sua visão a respeito dessa classificação. Não obstante, o construtivista relativista, embora não concorde com a formação positivista, consegue aceitá-la como uma dentre várias construções possíveis. O construtivista pode considerar a visão positivista malformada e rudimentar, mas não *errada* ou *desleal*.

Não estamos afirmando que este livro tem uma resposta “real” ou “definitiva” para as perguntas que levantamos, tanto sobre o que é avaliação quanto sobre como ela deve ser conduzida. Enxergamos nosso trabalho apenas como outra construção. Esperamos que o leitor considere este livro razoavelmente fundamentado e esclarecido, mas ele certamente está longe de ser uma verdade universal. Aliás, *não existe nenhuma verdade universal* em relação à qual nossa construção guarde mais ou menos uma boa semelhança. Acreditamos que um diálogo dialético contínuo sobre o que temos a dizer abrirá caminhos para *reconstruções* de maior poder e valor, mas não mais verdadeiras. Em última análise, todas essas construções, aquelas que temos no momento e aquelas que ainda teremos, são construções da *mente humana* e, por isso, estão sujeitas a todos os tipos de erro e fraqueza que sempre atormentam os empreendimentos humanos. Se existe um imperativo moral incorporado em nosso trabalho, esse imperativo é simplesmente este, de que continuaremos a nos esforçar ao máximo para buscar e considerar todas as reações e críticas que pudermos e de que tentaremos lidar com elas, mesmo que isso signifique abandonar totalmente nossa construção presente e abraçar uma em tudo diferente (e isso para nós encerra, justamente, o significado da expressão “revolução de paradigma”).

Com esse intuito, acolhemos possíveis trocas significativas com qualquer pessoa, independentemente de seu atual sistema de crenças, *contanto que*

ela esteja comprometida com um princípio hermenêutico-dialético que abra sua mente para outras construções e se disponha a mudar à proporção que seu conhecimento e esclarecimento aumentarem e a seguir qualquer direção que venha a se evidenciar nessa troca como algo frutífero. Quanto a nós, assumimos o compromisso de ter essa mesma postura.

A nosso ver, não é possível haver uma acomodação entre o sistema de crenças positivista e o construtivista do modo como eles são formulados no momento. Não vemos nenhuma possibilidade de acomodação, se esse ajuste implicar que um paradigma deva suplantar o outro pelo poder absoluto de seus argumentos ou que os paradigmas devam desempenhar funções complementares, ou ainda que um é apenas um caso especial do outro. *Na verdade, acreditamos* que, por meio do processo hermenêutico-dialético, surgirá uma nova construção nem “melhor” nem “mais verdadeira” que suas predecessoras, mas simplesmente mais fundamentada e esclarecida. Essa troca significa, em nossa opinião, a representação mais perfeita da frase “mercado acadêmico de ideias”.

Com esse intuito, esperamos que ninguém leia este livro com a intenção de encontrar uma “fórmula” para conduzir uma avaliação. Não é nosso desejo que ninguém conclua que agora, finalmente, sabemos o que fazer e de que forma fazer. Como bem ressaltou Mary Hesse (1980), do mesmo modo que todas as teorias científicas mais dia menos dia se provam falsas, o mesmo ocorrerá com todas as teorias que hoje nutrimos. Não estamos propondo uma nova ortodoxia e não gostaríamos que os leitores usassem nosso livro para essa finalidade.

Tomemos um momento para descrever brevemente o conteúdo deste livro. O capítulo 1 apresenta uma visão histórica do desenvolvimento da avaliação por meio do que concebemos como suas três primeiras gerações, analisa as principais críticas dirigidas a essas primeiras abordagens e defende que a avaliação responsiva, do modo como foi proposta por Robert Stake, supera a maioria desses problemas. A avaliação de quarta geração é apresentada como um modelo que pode incorporar proveitosamente essas técnicas responsivas.

O Capítulo 2 é dedicado à discussão sobre a necessidade de um novo paradigma. Os problemas da filosofia positivista são examinados brevemente e o paradigma construtivista é proposto como uma alternativa competente. O capítulo 3 analisa detalhadamente esse paradigma, contrasta

seu sistema de crenças básico com o do positivismo e apresenta argumentos e evidências que apoiam essa recusa ao positivismo e a adoção do construtivismo em seu lugar. Inúmeros caminhos propostos para alcançar essa acomodação entre os dois paradigmas foram analisados e rejeitados.

O capítulo 4 dedica-se a questões relacionadas ao papel da ética e da política na avaliação. Os problemas éticos e políticos são examinados do ponto de vista de ambos os paradigmas, concluindo-se que, embora ambos estejam sujeitos a dificuldades nesses âmbitos, o paradigma construtivista oferece várias vantagens, particularmente a concessão de poder (*empowerment*) e direitos a todos os grupos de interesse.

Já o capítulo 5 examina detalhadamente o processo hermenêutico-dialético, a princípio com uma discussão a respeito das construções enquanto realidades criadas, e procura investigar de que modo essas construções são formadas, contestadas, mudadas e conciliadas. As condições para um processo hermenêutico-dialético produtivo são apresentadas e as etapas processuais são examinadas em detalhe.

O capítulo 6 investiga a questão sobre se os paradigmas influenciam as metodologias. As metodologias, ou as estratégias gerais para a utilização de instrumentos e técnicas (métodos), com frequência são confundidas com os métodos em si; é por esse motivo, por exemplo, que o paradigma construtivista algumas vezes é chamado de paradigma “qualitativo”, como se tivesse sido definido totalmente com base em uma predileção por métodos qualitativos. Esse capítulo faz a distinção entre a “avaliação naturalista (construtivista) do primeiro tipo”, que não é outra coisa senão a avaliação qualitativa dentro do paradigma positivista, e a “avaliação naturalista (construtivista) do segundo tipo”, que se origina do sistema de crenças básico do construtivismo e, por isso, requer uma metodologia em grande medida distinta. As metodologias de ambos os paradigmas, convencional e construtivista, são investigadas mais detalhadamente a fim de esclarecer melhor essa distinção.

O capítulo 7 delinea a metodologia preferencial para a condução da avaliação de quarta geração. O leitor é conduzido por todas as etapas de implementação: do contrato inicial, da identificação dos grupos de interesse e suas várias reivindicações, preocupações e questões, da troca hermenêutico-dialética, na qual as construções de cada grupo são expostas a críticas e reações de todos os outros grupos, da elaboração de uma agenda

de negociação e do processo para discuti-la, até a divulgação final dos resultados.

O capítulo 8 examina a questão dos critérios de adequação ou qualidade. Se quisermos que as avaliações desse novo modelo sejam consideradas seriamente investigações disciplinadas, é indispensável demonstrar que elas atendem a determinados critérios. Evidentemente, o problema é especificar quais critérios são *apropriados* à avaliação de quarta geração. São propostos dois conjuntos de critérios: os assim chamados critérios de *fidedignidade*, que desenvolvemos como equivalentes paralelos àqueles normalmente especificados nas investigações positivistas (por isso com frequência são chamados de critérios "paralelos" ou "equivalentes"), e os assim chamados critérios de *autenticidade*, que foram criados com base no sistema de crenças básico do próprio paradigma construtivista. Enfatiza-se também o processo hermenêutico-dialético como um processo de controle de qualidade em si.

Por fim, o capítulo 8 analisa e destaca as consequências da adoção do modelo de avaliação de quarta geração: quanto às consequências para a *avaliação enquanto processo*, por exemplo, tem-se um processo sociopolítico contínuo, recorrente e divergente; quanto às consequências para a *função da avaliação*, por exemplo, tem-se uma mudança da função de técnico, descritor e/ou julgador para a de colaborador e negociador; e quanto às consequências para os *resultados da avaliação*, por exemplo, examinam-se determinados princípios que a avaliação de quarta geração talvez queira seguir na condução desse tipo emergente de avaliação.

AVALIAÇÃO: ATINGINDO A MAIORIDADE*

Como o objetivo deste livro é falar sobre *avaliação*, parece razoável iniciá-lo com uma definição que esclareça o que queremos dizer com esse termo. Contudo, propor uma definição neste momento, além de arbitrário e precoce, seria contraproducente aos temas centrais deste livro. Isso porque, para nós, não existe *nenhuma* forma “correta” de definir *avaliação*, pois, se fosse possível encontrar esse sentido, isso poria fim, de uma vez por todas, à discussão acerca de como a avaliação deve ser conduzida e sobre quais são seus propósitos. Consideramos as definições de avaliação interpretações mentais humanas, cuja correspondência com determinada “realidade” *não é e não deve* ser um problema. Não existe resposta para a pergunta “Mas o que de fato é avaliação” e não faz sentido fazê-la.

Em vez disso, começaremos por esboçar brevemente os vários significados atribuídos à avaliação nos últimos cem anos, designações que refletem o contexto histórico existente, os propósitos que as pessoas tinham em mente quanto à condução das avaliações e as suposições filosóficas que os avaliadores, teóricos e também outros profissionais estavam inclinados a fazer. Defenderemos que, com o passar do tempo, a construção da avaliação se tornou mais fundamentada e esclarecida, a ponto de, no presente

* Grande parte do conteúdo deste capítulo é extraída de nosso artigo “The countenances of fourth generation evaluation — Description, judgment, and negotiation” (in *Evaluation studies review annual*, Newbury Park, Califórnia, Sage, 1986, vol. 1, pp. 70-80). Uma versão ligeiramente diferente encontra-se em *The politics of evaluation* (Dennis J. Palumbo, ed., Newbury Park, Califórnia, 1988).

momento, ser possível conceber uma construção radicalmente nova, a que chamamos de avaliação de quarta geração. Obviamente, não há nenhum consenso a respeito dessa forma de avaliação, visto que não houve consenso também sobre os modelos anteriores — uma situação que uma breve passada de olhos em qualquer texto convencional sobre avaliação rapidamente confirma. Todavia, nós a apresentamos como uma construção que, acreditamos, contenha, ou ao menos atenua, as imperfeições, as lacunas e a ingenuidade das formulações precedentes.

Não acreditamos que tenhamos encontrado a formulação essencialmente correta. Entretanto, *estamos* preparados para defender que a construção a que chamamos de avaliação de quarta geração é mais bem fundamentada e esclarecida do que as construções anteriores. No entanto, do mesmo modo que esses modelos precedentes, a avaliação de quarta geração mais cedo ou mais tarde também se provará de algum modo inadequada e exigirá reforma, refinamento, ampliação e, provavelmente, total substituição. Aliás, acreditamos ser nosso dever identificar aspectos da avaliação para os quais esse modelo *não* é adequado, rumo a um esforço contínuo de *reconstrução*.

Diante do exposto, podemos começar.

A primeira geração: mensuração

A avaliação tal como a conhecemos não apareceu simplesmente do dia para a noite; é uma consequência de um processo de evolução de construção e reconstrução que envolve inúmeras influências interatuantes.

Dentre as influências iniciais, a mais importante é a mensuração de diferentes atributos de crianças e jovens em idade escolar. Os exames escolares foram utilizados durante centenas de anos para avaliar se os alunos haviam conseguido “dominar” o conteúdo de diversos cursos ou assuntos nos quais haviam sido iniciados. Conteúdos apropriados eram definidos com base em sua autoridade, fossem eles sobre Aristóteles, a Bíblia ou, mais recentemente, as descobertas da ciência. O principal objetivo da escola era ensinar às crianças *o que se reconhecia como certo*; para demonstrar que tinham domínio, as crianças tinham de regurgitar esses “fatos” em exames, que eram em essência testes de memória. Os primeiros exames

escolares eram administrados sobretudo oralmente, um aluno por vez, e exigiam, se fossem por escrito, respostas “dissertativas”.

Portanto, não é de surpreender que o primeiro exemplo de pesquisa educacional publicado, “The futility of the spelling grind” [A futilidade da labuta da ortografia] (Rice, 1897), tenha dependido de pontuações de teste para compor seus dados. Joseph Mayer Rice estava alarmado com o fato de *todo* o tempo escolar ser devotado, pelo menos nas escolas americanas, ao que hoje em dia chamaríamos de “conceitos básicos”. Rice achava que, se fosse possível tornar as escolas mais competentes, isto é, se esse mesmo aprendizado básico pudesse ocorrer em *menos* tempo, o currículo escolar poderia ser ampliado para incluir arte, música e outros temas que ele desejava ver incluídos (o “supérfluo”). Depois de uma iniciativa fracassada de combater o problema do currículo de forma geral (malsucedida porque ele não conseguiu estabelecer controles de campo adequados para um empreendimento tão abrangente), Rice concentrou-se no exemplo clássico da ortografia. Ele concebeu um teste de ortografia que ele mesmo ministrava em inúmeras escolas amplamente dispersas em termos geográficos, coletando em cada escola dados sobre a quantidade de tempo dedicado ao ensino de ortografia. Sua análise subsequente indicou que não havia praticamente nenhuma relação entre o tempo gasto no estudo de ortografia e o êxito posterior no teste. As notas obtidas foram consideradas provas concretas do grau de aproveitamento dos alunos.

Outra aplicação de testes que viria a ter grande importância ocorreu na França. O ministro da Educação francês, perseguido pelos professores que exigiam que ele encontrasse alguma solução para fazer a triagem dos jovens com retardamento mental, os quais, como se afirmou, estavam inviabilizando o ensino a crianças “normais” (um tipo de integração reversa, por assim dizer), pediu ao psicólogo Alfred Binet que idealizasse um teste para essa finalidade. Primeiramente, Binet tentou utilizar técnicas psicométricas que haviam sido aperfeiçoadas na Inglaterra por Francis Galton (antes disso, um membro de sua equipe, Karl Pearson, havia criado o coeficiente de correlação produto-momento para análise de dados) e, na Alemanha, por Wilhelm Wundt. Quando essas técnicas se demonstraram ineficientes, Binet idealizou uma nova abordagem, baseada na observação de senso comum de que os jovens com retardamento mental não seriam capazes de lidar com situações simples da vida, como contar dinheiro ou identificar

utensílios domésticos, nem com seus semelhantes normais. No final, Binet acabou conseguindo organizar as tarefas de acordo com a idade dos sujeitos do estudo normalmente capazes de concluí-las e, além disso, cunhar o termo “idade mental”. Por volta de 1912, já era comum dividir a idade mental obtida pela idade cronológica do sujeito do estudo para determinar o “quociente de inteligência”. O teste de Binet transpôs rapidamente o Atlântico em 1910, quando foi traduzido por Henry Golddard; em 1916, quando Louis Terman reformou e renormalizou o teste de Binet (agora denominado Stanford-Binet) para aplicá-lo em crianças americanas, o teste de QI tornou-se parte permanente do sistema americano.

A utilidade dos testes para propósitos escolares foi bem recebida pelos profissionais na liderança. Em 1904, a Associação Americana de Educação (National Education Association — NEA) designou um comitê para que investigasse a possibilidade de utilizar esses testes na classificação e determinação do aproveitamento escolar das crianças; essa associação nomeou três outros comitês em 1911. Em 1912, foi criado o primeiro departamento de pesquisa de distrito escolar na cidade de Nova York; sua responsabilidade era realizar um levantamento contínuo do sistema, usando as “novas técnicas de mensuração”. Departamentos semelhantes logo em seguida foram criados em outras cidades importantes; os diretores desses departamentos, que normalmente ocupavam o cargo de superintendente assistente, começaram a reunir-se anualmente com a Associação Americana de Administradores Escolares; em seguida, eles se organizaram de uma maneira mais formal, transformando-se na Associação Americana de Pesquisa Educacional.

Provavelmente, a influência mais importante, responsável pelo rápido avanço e pela aceitação dos testes mentais, foi a necessidade de fazer a triagem de pessoal para alistamento nas forças armadas, na Primeira Guerra Mundial. Os comandantes militares contaram com o apoio da Associação Americana de Psicologia (American Psychological Association — APA) para idealizar um instrumento apropriado. A APA designou um comitê, presidido por Arthur Otis, para empreender essa tarefa, que foi concretizada em tempo recorde e com excelência. O primeiro teste de inteligência aplicado a *grupos*, o Army Alpha, foi ministrado com êxito em mais de 2 milhões de homens. Incitado por esse sucesso, Otis comprometeu-se a alterar e adaptar esse instrumento para ser aplicado em escolas.

Vários fatores que pareciam apenas indiretamente relacionados ao teste estavam também destinados a desempenhar funções importantes no desenvolvimento dessa primeira geração de avaliação. O primeiro desses fatores contextuais foi a legitimação propiciada pela fenomenal ascensão das ciências *sociais*. Quando John Stuart Mill recomendou em 1843 a aplicação da “abordagem científica” à pesquisa de fenômenos humanos/sociais, recomendação fundamentada no enorme sucesso dessa abordagem na física e na química do final do século XVIII e início do século XIX e na ausência de uma base sistemática para pesquisas em ciências “humanas”, ele mal podia prever o entusiasmo com que sua sugestão seria acolhida ou as consequências de amplo alcance que a adoção dessa abordagem teria. As primeiras tentativas de importância para torná-la “científica” foram incitadas pela tese de Darwin de que até mesmo as pequenas diferenças na estrutura animal ou vegetal poderiam, quando acumuladas por longos períodos, ter consequências funcionais extremamente significativas para as espécies. Se assim for, começaram a inferir os cientistas sociais, então pequenas diferenças nas pessoas talvez também sejam essenciais para compreender padrões de desenvolvimento importantes nos seres humanos. Foi por esse motivo, dentre outros, que os laboratórios psicométricos já mencionados foram criados por Galton, em 1873, e por Wundt, em 1879. Determinadas descobertas, como a de que diferenças individuais no tempo de resposta eram características de sujeitos humanos, levavam a crer que esses investigadores estavam no caminho certo. Particularmente a psicologia passou a dedicar-se a essa nova abordagem científica, tentando igualar-se ao máximo às ciências físicas. Como seria de esperar, esse intento foi bem servido pelas mensurações aparentemente precisas que os testes estavam propiciando. Quando, em meados da década de 1920, Ronald Fischer, trabalhando como estatístico para o setor inglês de algodão, criou instrumentos analíticos básicos, bem como as tabelas matemáticas necessárias para interpretar seus resultados, as ciências sociais, dentre as quais a da educação, já estavam seguindo de perto as pegadas de suas tão admiradas congêneres, as ciências duras ou experimentais (*hard sciences*).

O segundo fator contextual que estimulou a aplicação de testes foi o surgimento do movimento da administração científica nos setores empresarial e industrial. Se os seres humanos são o principal elemento na produção de bens e serviços, a responsabilidade do administrador é tornar

seu trabalho o mais eficiente e eficaz possível. Esse movimento, que se principiou antes da Primeira Guerra Mundial, mas floresceu na década de 1920, fundamentou-se expressivamente em pesquisas sobre tempo e movimento para determinar os métodos de trabalho mais produtivos e em taxas salariais de trabalho por empreitada para que os trabalhadores se sentissem motivados a se submeter a uma disciplina tão fatigante e não gratificante em termos pessoais quanto essa. Na época em que essa visão foi criteriosamente contestada pelos Estudos de Hawthorne (Roethlisberger e Dixon, 1939), o etos da administração científica já havia se introduzido nas escolas. Os alunos eram vistos como "matéria-prima" a ser "processada" na "fábrica" da escola, presidida, de maneira suficientemente apropriada, pelo "superintendente" escolar. Os testes desempenhavam uma função decisiva para o desdobramento dessa representação; eles eram vistos como um meio para determinar se os alunos estavam à altura das "especificações" que a escola havia estabelecido, especificações que, a essa época da história, eram principalmente e intrinsecamente preparatórias para a faculdade.

Todas essas influências acabaram provocando uma notável proliferação dos testes escolares durante as décadas de 1920 e 1930. Os testes de aproveitamento, criados em 1908 com a publicação do Stone reasoning test in arithmetic [Teste de Stone de raciocínio em aritmética], passaram a ser produzidos de várias formas. A publicação em 1922 do Stanford achievement battery [Bateria de testes de aproveitamento de Stanford] ofereceu um instrumento que, pela primeira vez, possibilitou a avaliação simultânea da situação relativa de um aluno em relação a várias disciplinas escolares. Gertrude Hildreth, da Universidade de Minnesota, comprometeu-se a publicar uma lista de testes suficientemente bons para garantir sua adoção; sua bibliografia de testes mentais e escalas de classificação, de 1933, continha mais de 3,5 mil títulos. Sua terceira e última bibliografia, lançada em 1945, continha mais de 5,2 mil. E, durante esse período, os termos *mensuração* e *avaliação* passaram a ser usados intercambiavelmente. Os livros sobre esses assuntos em geral traziam ambos os termos nos títulos — embora o termo *mensuração* viesse invariavelmente em primeiro lugar.

Portanto, a primeira geração de avaliação pode legitimamente ser chamada de *geração da mensuração*. A função do avaliador era técnica; ele devia conhecer o arsenal de instrumentos disponíveis para que, desse modo,

qualquer variável que se desejasse investigar pudesse ser mensurada. Se não houvesse nenhum instrumento apropriado, supunha-se que o avaliador tivesse o conhecimento necessário para criá-lo. Além disso, é extremamente importante mencionar, essa primeira geração ou acepção técnica de avaliação persiste ainda hoje, como se pode comprovar, por exemplo, pela prática frequente de exigir que os alunos passem nos testes, como parte dos procedimentos de graduação na escola de ensino médio ou de admissão nas faculdades, pela aplicação desses testes em vários estados americanos, para classificar as escolas e mesmo a eficácia de cada um dos professores, e pela sucessiva publicação de textos que usam o termo *mensuração e avaliação* nos respectivos títulos (consulte, por exemplo, Gronlund, 1985).

A segunda geração: descrição

A segunda geração de avaliação ganhou existência em virtude de uma séria deficiência na primeira geração, isto é, escolher como sujeitos* da avaliação os *alunos*. Isso porque, logo após a Primeira Guerra Mundial, tornou-se evidente que os currículos escolares precisavam passar por uma drástica reforma e que uma abordagem de avaliação que não conseguia oferecer outra coisa senão dados sobre os alunos não poderia servir aos propósitos da avaliação que naquele momento era contemplada.

Logo após a guerra, as escolas secundárias americanas começaram a testemunhar uma afluência de estudantes que raramente haviam ultrapassado o nível elementar em oportunidades anteriores. Esses alunos exibiam necessidades e aspirações que não podiam ser atendidas de modo adequado pelos currículos preparatórios para faculdade então prevalentes. Muitos deles viam a escola secundária como oportunidade de adquirir as habilidades essenciais para suplantarem a posição social e econômica de seus pais, mas as escolas estavam malpreparadas para oferecer tal ensino. Além disso, as iniciativas de idealização de currículos mais apropriados foram

* Como será possível comprovar no capítulo 2, os autores deste livro hesitam em empregar o termo objeto em relação a seres humanos. O termo original utilizado é *evaluand*, o qual, em respeito à preferência dos autores, optamos por traduzir por *sujeito-objeto* da avaliação nos casos em que não há especificação se esse "objeto" é ou não humano. (N. da T.)

Avaliação

frustradas antes mesmo que eles fossem submetidos a uma avaliação justa, na medida em que as escolas secundárias estavam inextricavelmente presas ao sistema de unidade Carnegie (que especificava os tipos e a quantidade de unidades de crédito necessários para colar grau). Os principais obstáculos na trajetória para alterar essa exigência foram, não surpreendentemente, as faculdades e universidades, que temiam que, se o sistema Carnegie fosse abolido enquanto princípio para a acumulação de créditos na escola secundária, elas seriam forçadas a aceitar os alunos graduados na escola secundária malpreparados para enfrentar os currículos universitários convencionais.

O objetivo do programa Eight-Year Study (Estudo de Oito Anos), introduzido em 1933, era determinar a validade desse ponto de vista. Trinta escolas secundárias públicas e particulares receberam autorização para desenvolver currículos mais responsivos, entendendo que seus alunos graduados seriam admitidos em faculdades cooperantes sem ter, necessariamente, de atender às exigências da Fundação Carnegie. O propósito do programa Eight-Year Study era demonstrar que os alunos capacitados de acordo com esses currículos não convencionais mesmo assim seriam capazes de ter um bom desempenho na faculdade. O período de oito anos foi escolhido para possibilitar que pelo menos uma coorte desses alunos concluísse quatro anos completos da escola secundária e quatro anos do ensino superior.

Um dos problemas imediatos enfrentados pelos idealizadores desse estudo foi criar uma forma de avaliar se o desenvolvimento dos novos currículos estava ocorrendo de acordo com o planejado. Afinal de contas, esse teste não seria justo se os alunos fracassassem na faculdade não porque os currículos fossem *em princípio* inadequados, mas apenas porque o eram *na prática*. Por uma feliz coincidência, Ralph W. Tyler, membro do Departamento de Pesquisa Educational, na Universidade Estadual de Ohio¹, *campus* em que o programa Eight-Year Study foi implementado, havia trabalhado durante vários anos com determinadas faculdades dessa universidade com a intenção de desenvolver testes para avaliar se os alunos

¹ Por uma curiosa virada do destino, um de nós teve o privilégio de dirigir esse mesmo departamento (que antes se chamava Departamento de Pesquisas e Serviços Educacionais) de 1961 a 1965.

havam ou não aprendido o que seus professores *pretendiam que eles aprendessem*. Esses resultados desejados de aprendizagem foram chamados de *objetivos*. Tyler estava comprometido em realizar o mesmo tipo de trabalho com as escolas secundárias do programa Eight-Year Study, mas havia uma alteração importante em relação à avaliação convencional (mensuração): a finalidade dos estudos seria *aperfeiçoar os currículos que estavam sendo desenvolvidos e confirmar se eles estavam funcionando*. Foi assim que nasceu a avaliação de programas.

Assim que as escolas participantes desse programa começaram a idealizar seus novos currículos, Tyler coletou informações sobre o nível de aproveitamento dos alunos nos programas em relação aos objetivos que haviam sido definidos. Essas informações, paralelamente a uma análise dos pontos fortes e fracos que consequentemente ficaram aparentes, foram utilizadas para conduzir aprimoramentos e reformas — um processo que hoje chamaríamos de avaliação *formativa*, com a exceção de que os resultados só podiam ser obtidos *depois*, e não *durante* um determinado teste. Esse processo repetia-se sucessivamente a cada curso oferecido, até que se concluísse que o currículo estava produzindo um nível apropriado de aproveitamento.

Portanto, foi assim que surgiu o que hoje optamos por chamar de avaliação de segunda geração, uma abordagem caracterizada pela *descrição* de padrões de pontos fortes e fracos com respeito a determinados objetivos estabelecidos². O avaliador cumpria a função de *descriptor*, embora também se mantivessem os aspectos técnicos anteriores dessa função. A mensuração então deixou de ser tratada como equivalente à avaliação, sendo porém redefinida como *um dos vários instrumentos* que poderiam ser empregados em seu serviço. Quando o relatório do Eight-Year Study foi publicado em 1942, o terceiro volume, que descrevia as atividades de avaliação do projeto (Smith e Tyler, 1942), atraiu grande atenção. Tal como Lord Byron, Tyler ficou famoso do dia para a noite. Posteriormente, seria reconhecido como o “Pai da Avaliação” (Comitê Misto, 1981).

2 Escolhemos o termo *descrição* para caracterizar essa geração porque é o termo usado no agora clássico artigo de Robert Stake (1967). Essa geração poderia ter sido chamada igualmente bem de *geração de objetivos*.

A terceira geração: juízo de valor

A abordagem descritiva centrada em objetivos tinha algumas falhas graves, embora, até o período pós-Sputnik (1957), elas não fossem tão perceptíveis. Foi nessa época que essa abordagem se provou inadequada à missão de avaliar a resposta do governo federal às supostas deficiências da educação americana que haviam permitido que os russos ganhassem vantagem na exploração do espaço: os programas de aperfeiçoamento do conteúdo do curso da Fundação Nacional da Ciência (National Science Foundation — NSF) (BSCS Biologia, Projeto CHEM, PSSC Física e SMSG Matemática)* e da então Secretaria de Educação (Projeto de Inglês e Projetos de Estudos Sociais). Quando os avaliadores indicados para essas equipes de projeto insistiram em que não poderiam iniciar o trabalho enquanto não tivessem à mão os objetivos do projeto, eles foram dispensados pelos criadores do programa (os quais, é necessário lembrar, eram físicos, químicos, biólogos e matemáticos profissionais nos projetos iniciais da NSF, e não professores de ciências nem professores de matemática), que os consideraram inapropriados. Esses desenvolvedores não queriam comprometer-se com nenhum objetivo enquanto não tivessem uma visão mais clara do que estavam fazendo, e não queriam nem mesmo estabelecer objetivos provisórios para evitar que mais tarde percebam que isso teria castrado prematuramente sua criatividade. Além disso, eles não podiam aceitar uma estratégia de avaliação que não produzisse resultados enquanto o programa não estivesse totalmente desenvolvido; se, nesse caso, uma avaliação apresentasse deficiências, seria em vários aspectos tarde demais para tomar providências a respeito (tendo em vista em especial a atmosfera de crise, de nível nacional, na qual eles trabalhavam). Esses problemas estão bem documentados no artigo agora clássico “Course improvement through evaluation” [Aprimoramento de currículos por meio da avaliação] (1963), de Cronbach.

Contudo, os dissidentes sobre o que a essa época havia se tornado o método de avaliação aceito tinham outra crítica mais importante a dirigir. Visto que, por natureza, a avaliação de segunda geração tyleriana era es-

* Biological Science Curriculum Study (BSCS), CHEM (química), Physical Science Curriculum Study (PSCS), School Mathematics Study Group (MSG). (N. da T.)

sencialmente descritiva, ela desprezou o que Robert Stake chamou de *outra* fisionomia (*countenance*) ou face da avaliação em seu artigo de 1967 já citado anteriormente: o *juízo de valor*. Segundo Stake:

A fisionomia da avaliação contemplada pelo educador não é a mesma contemplada pelo especialista em avaliação. O especialista se vê como um “descritor”, alguém que descreve aptidões e ambiente e concretizações. Entretanto, o professor e o gestor escolar esperam que o avaliador classifique algo ou alguém pelo mérito. Além disso, esperam que o avaliador julgue as coisas com base em padrões externos, em critérios talvez pouco relacionados com os recursos locais da escola. Nenhum dos dois vê a avaliação de uma maneira suficientemente ampla. Tanto a descrição quanto o juízo de valor são essenciais — na verdade, ambos são procedimentos básicos de avaliação. (Citado em Worthen e Sanders, 1973, p. 109)

O apelo para incluir o juízo de valor no procedimento de avaliação foi o marco do surgimento da avaliação de terceira geração, uma geração na qual a avaliação se caracterizou por iniciativas que visavam alcançar *juízos de valor* e na qual o avaliador assumiu o papel de *jugador*, mantendo igualmente as funções técnicas e descritivas anteriores. Esse apelo, amplamente difundido entre os profissionais da área, em especial por Michael Scriven (1967), evidenciou vários problemas com os quais as gerações anteriores não haviam lidado adequadamente. Primeiro, era necessário considerar os *próprios* objetivos como algo problemático; os objetivos, tanto quanto o desempenho, deveriam ser submetidos a avaliação. Como bem assinalou um determinado gracejador, uma coisa que de forma alguma vale a pena fazer certamente não vale a pena fazer bem feito. Além disso, tal como Stake evidenciou, o juízo de valor depende da existência de *critérios* para que possa ser feito com base nesses critérios. Entretanto, a inclusão de critérios que devessem, por definição, ser carregados de valor em uma iniciativa científica presumidamente isenta de valor como a avaliação era repugnante para a maioria dos avaliadores. Em conclusão, para haver juízo de valor é necessário um julgador. Os avaliadores não se sentiam competentes para atuar nessa posição, achavam presunçoso fazê-lo e temiam a vulnerabilidade política à qual se expunham em virtude disso. Não obstante, eram incitados a aceitar essa responsabilidade, em grande medida

pelo fato de que, dentre todos os possíveis candidatos a julgador, os avaliadores eram sem dúvida os mais objetivos (Scriven, 1967).

Em última análise, o apelo ao juízo de valor não podia ser ignorado e os avaliadores em pouco tempo conseguiram enfrentar o desafio. Uma profusão de novos modelos de avaliação surgiu em 1967 e depois disso: os modelos neotylerianos que abrangiam o próprio Modelo de Fisionomia de Stake (1967) e o Modelo de Avaliação da Discrepância (Provus, 1971); modelos orientados à tomada de decisões, como Contexto, Insumo, Processo e Produto (Context, Input, Process e Product — CIPP) (Stufflebeam et al., 1971); modelos orientados a resultados, como o Modelo sem Referência a Objetivos (Scriven, 1973); modelos de neomensuração sob o pretexto de experimentação social (Boruch, 1974; Campbell, 1969; Rivlin e Timpane, 1975; Rossi e Williams, 1972); e modelos genuinamente críticos, como o Modelo de Crítica Artística (Eisner, 1979). Contudo, todos esses modelos pós-1967 tinham a mesma visão com relação a uma questão: o juízo de valor era parte essencial da avaliação. Todos preconizavam, de uma maneira mais ou menos explícita, que o avaliador fosse um julgador (especialista). Havia diferenças com respeito ao grau segundo os avaliadores eram representados como julgadores adequados, desde a indecisão dos modelos orientados à tomada de decisões — cujos proponentes hesitavam em defender uma função crítica agressiva porque isso, ao que parecia, cooptava os próprios tomadores de decisões para os quais as avaliações supostamente serviriam —, aos defensores do Modelo de Avaliação de Discrepância — que se autoatribuíam a função de ajudar o cliente a determinar critérios de juízo de valor —, à positividade dos defensores dos modelos de juízo de valor nos quais o avaliador era escolhido precisamente pelas qualidades de sua especialidade (*connoisseurship*). Não obstante, parece justo afirmar que durante essa década, e mais intensamente depois de 1967, o juízo de valor tornou-se uma marca distintiva dos avaliadores da terceira geração.

Problemas predominantes das três primeiras gerações

Embora a discussão precedente sobre as três primeiras gerações de avaliação tenha sido breve, ela é suficiente para demonstrar que cada geração

subsequente representou um passo adiante, tanto em relação à variedade de sentidos ou conteúdos incluídos na construção defendida quanto em seu nível de sofisticação. A coleta de dados sistemática sobre indivíduos só se tornou possível com o desenvolvimento de instrumentos apropriados, do tipo que caracterizou a primeira geração. Entretanto, a avaliação teria estagnado nesse nível se a segunda geração não tivesse apresentado uma solução para avaliar também os vários objetos de avaliação (*evaluands*) não humanos — os programas, os conteúdos, as estratégias de ensino, os padrões organizacionais e as “abordagens” em geral. A terceira geração exigiu que a avaliação se centrasse no juízo de valor, tanto sobre o mérito do objeto de avaliação — seu valor íntimo e intrínseco — quanto sobre sua importância — seu valor extrínseco ou contextual (Guba e Lincoln, 1981). Porém, todas as três gerações, enquanto grupo, padeceram e continuam padecendo de algumas imperfeições ou defeitos suficientemente graves para justificar que se questione se não seriam indispensáveis no momento outros aprimoramentos — ou mesmo uma completa reconstrução. Acreditamos que haja pelo menos três imperfeições ou defeitos mais sérios: tendência ao gerencialismo, incapacidade de acomodar o pluralismo de valores e comprometimento exagerado com o paradigma científico de investigação. A seguir, examinaremos brevemente cada uma delas; ao longo do livro, retornaremos a esses temas com frequência.

Tendência ao gerencialismo. O termo *gestor/administrador* inclui uma variedade de indivíduos, mas na maioria das vezes representará clientes ou patrocinadores que contratarão ou financiarão uma avaliação, bem como os profissionais de liderança aos quais os representantes responsáveis pela implementação do objeto de avaliação (e não da avaliação em si) se reportam. Essa última categoria abrange, por exemplo, os membros do conselho escolar, o superintendente e os diretores (às vezes) das escolas em que uma inovação curricular está sendo experimentada; o administrador hospitalar e o chefe de serviços de enfermagem de um hospital no qual novos métodos de tratamento de pacientes oncológicos estão sendo instituídos; ou o diretor de uma instituição social na qual novos programas destinados a oferecer oportunidades de lazer para deficientes físicos estão sendo comparados. São os administradores com os quais o avaliador normalmente firma um acordo de avaliação, aos quais ele acata no estabelecimento de parâmetros e delimitações e aos quais ele se reporta. Essa relação tra-

dicional entre administradores e avaliadores raras vezes é contestada; não obstante, ela gera inúmeras consequências extremamente indesejáveis.

Primeiro, em vista dessa estrutura, o administrador é efetivamente *eximido de responsabilidade*. Na medida em que o administrador *fica fora* da avaliação, suas qualidades e os métodos administrativos que emprega não podem ser questionados e tampouco ele pode ser responsabilizado pelo que o objeto de avaliação produz ou não. Se houver alguma falha, a avaliação necessariamente apontará o dedo para outro lugar.

Segundo, a relação característica entre administrador e avaliador é *enfraquecedora* (privadora de poder) e *injusta*. O administrador tem o poder supremo de determinar as perguntas que a avaliação tentará responder, a forma como as respostas serão coletadas e interpretadas e a quem os resultados serão divulgados. Obviamente, essas questões são muitas vezes decididas após prévia consulta ao avaliador. Entretanto, em caso de desacordo, a decisão final é do administrador. O único recurso que o avaliador tem é recusar-se a conduzir a avaliação em condições que não lhe agradem. Essa circunstância na realidade tira o poder dos grupos de interesse ou interessados (*stakeholders*) que talvez queiram obter respostas para outras perguntas e tenham outras formas de respondê-las e outras interpretações a respeito. É difícil, quando não impossível, conduzir uma avaliação por meio de qualquer um dos métodos das três primeiras gerações mantendo-se aberto a recomendações (*inputs*) de outros grupos de interesse. O processo como um todo é evidentemente injusto com esses outros grupos, cujas possíveis recomendações não são solicitadas nem apreciadas, ao passo que o administrador é erguido a uma posição de mais alto poder.

Terceiro, a relação característica entre administrador e avaliador é *privadora de direitos*. Na maioria das vezes, o administrador conserva o direito, contratualmente, de determinar se os resultados da avaliação devem ser disponibilizados e, em caso afirmativo, para quem. Não é incomum os avaliadores barganharem o poder de divulgar informações pelo direito de produzir qualquer relatório que julgarem adequado. Parecia uma permuta razoável: o avaliador protegia sua integridade ao conservar as prerrogativas editoriais, ao passo que o administrador, por sua vez, deliberava sobre questões de divulgação. Porém, os interessados que não eram informados sobre os resultados eram efetivamente impedidos de tomar qualquer tipo de medida que porventura esses resultados pudessem lhes recomen-

dar e, inclusive e sobretudo, de proteger seus próprios interesses. Eles eram privados do privilégio de obter informações e, com isso, de seus direitos.

Em conclusão, a relação característica entre administrador e avaliador tem grande propensão para se tornar *amistosa*. Conceder ao administrador o direito de determinar o formato de uma avaliação significa, na realidade, estar em conluio com ele. Obviamente, estabelecer esse conluio é vantajoso tanto para o administrador quanto para o avaliador. Em relação ao administrador, a avaliação conduzida de uma forma tal que o isenta de responsabilidade, tirando o poder e privando de direitos possíveis antagonistas, é evidentemente preferível à avaliação que o responsabiliza e possibilita que esses antagonistas assumam uma ínfima fração de poder. Quanto ao avaliador, a avaliação realizada de uma maneira tal que ganhe a aprovação do administrador tende a viabilizar novos contratos e garantir uma fonte de renda contínua. Henry M. Brickell uma vez mencionou que o avaliador forçosamente se envolve em um delicado malabarismo: “Morder a mão que o alimenta passando sempre a impressão de que está apenas lambendo”. Esse malabarismo fica mais fácil quando o avaliador decide não morder de forma alguma. Embora a vasta maioria dos avaliadores provavelmente hesitasse em participar de maneira deliberada desse conluio, ainda assim, em qualquer uma das três gerações, era demasiado fácil envolver-se nessa situação.

Michael Scriven escreveu amplamente sobre o problema do gerencialismo (1983); a solução que ele oferece para esse problema é o envolvimento com um modelo de avaliação que faça perguntas supostamente interessantes para o *consumidor* e que preste contas a esse grupo. Ele projeta avaliações equivalentes às análises encontradas na publicação *Consumer Reports*. Essa abordagem representa um grande avanço porque reconhece a importância de outros grupos além dos administradores. Porém, não há nenhum motivo convincente para *eliminar* os administradores apenas com o intuito de evitar a possibilidade de gerencialismo; há maneiras mais adequadas de lidar com esse problema, como demonstraremos. A orientação para o consumidor de fato agrega um público de interesse a esse *mix*; restam ainda outros, incluindo os administradores, que não estão incluídos na abordagem orientada ao consumidor.

Incapacidade de acomodar o pluralismo de valores. É comum acreditar que as sociedades compartilham valores, que existe um conjunto de valo-

res que caracteriza os membros de uma sociedade à qual todos os membros se aculturam e com a qual todos concordam. A ideia da "grande mistura de culturas" que absorve imigrantes e, de certo modo, os transforma em americanos é um exemplo. Outro é a expectativa de que as escolas transmitam "nossa tradição", uma frase que encerra a ideia de que nossa tradição é compartilhada. E um terceiro exemplo: normalmente se afirma que nosso sistema moral está fundamentado na "ética judaico-cristã".

Foi apenas nos últimos 20 anos que conseguimos compreender que essa sociedade, *nossa sociedade*, se caracteriza essencialmente pelo pluralismo de valores. Lições sobre pluralismo de valores foram elucidadas a todos nós no final da década de 1960, década que testemunhou não somente rivalidades tradicionais entre partidos políticos, mas também conflitos étnicos, de gênero e, infelizmente, até de gerações, os quais pareciam não ter solução.

Na avaliação, o apelo ao juízo de valor ocorreu, pela primeira vez, quase na mesma época em que houve certo reconhecimento do pluralismo de valores. Naturalmente, os valores já estavam implícitos na avaliação desde a primeira vez em que ela foi empregada; aliás, linguisticamente, o próprio termo *avaliação* está enraizado no termo *valor*. Contudo, era fácil ignorar o fato de que mesmo o desenvolvimento de um instrumento "objetivo" envolvia juízos de valor ou que a delineação de objetivos implicava um consenso quanto aos valores, contanto que a discussão sobre as diferenças valorativas não viesse à tona. Entretanto, uma vez levantada, essa discussão não podia ser empurrada de volta a seu invólucro. A dúvida sobre *quais valores* deveriam predominar em uma avaliação (isto é, os valores de quem) ou, de outro modo, de que forma as diferenças valorativas poderiam ser negociadas, no momento, evidencia-se como o problema mais importante.

Por muito tempo se defendeu que, não obstante a existência de diferenças valorativas, era possível confiar nos resultados de uma avaliação porque a metodologia usada é científica e a ciência é comprovadamente isenta de valores. A verdadeira intenção em exigir objetividade é obviar a questão da influência dos valores. Evidentemente, como se admitiu, o avaliador não tem nenhum controle sobre como os resultados da avaliação serão empregados; se as pessoas com valores diferentes optam por interpretar os resultados factuais de maneira distinta, dificilmente o avaliador pode ser responsabilizado. Contudo, como veremos, a alegação de que a ciência

é isenta de valores pode ser seriamente contestada. Se a ciência *não* for isenta de valores, então não apenas os resultados estarão sujeitos a interpretações diferentes, mas os próprios "fatos" serão determinados em interação com o sistema de valores em que o avaliador (provavelmente não de maneira intencional) se fundamenta. Portanto, *todo ato de avaliação se torna um ato político*. Aliás, todo ato de investigação, seja ele avaliação, pesquisa ou análise de políticas, torna-se um ato político nesse sentido.

Essa afirmação de isenção de valores em relação às avaliações é completamente evocativa (reforçadora) da tendência gerencialista já descrita. Se os valores não fazem nenhuma diferença, os resultados de uma avaliação representam então as circunstâncias *tal como elas de fato são*; elas devem ser aceitas como verdades objetivas. O fato de o administrador estabelecer os limites e os parâmetros do estudo seria então relativamente irrelevante, do mesmo modo como o seria o fato de ele controlar as perguntas levantadas, a metodologia e os resultados.

A pretensão de isenção de valores não é defensável, como demonstraremos. E sendo *isso* verdade, o pluralismo de valores de nossa sociedade é uma questão crucial a ser levada em consideração na avaliação. Sob hipótese alguma, nenhuma das abordagens de avaliação das primeiras três gerações leva em conta as diferenças valorativas.

Comprometimento exagerado com o paradigma científico de investigação. Como mencionamos, os cientistas sociais seguiram a recomendação de Mill de emular os métodos das ciências físicas com convicção e entusiasmo. Há diversos motivos para essa reação intensamente positiva, dentre eles o espetacular sucesso que as ciências físicas desfrutaram, o desejo dos cientistas sociais de serem racionais e sistemáticos, ao estilo de Descartes ("Penso, logo existo") e do positivismo em geral, e a necessidade de obter legitimidade enquanto profissão seguindo o mais rigorosamente possível a metodologia que caracterizou suas congêneres, as ciências duras ou experimentais.

As premissas do método científico eram em si cantos de sereia, pois pareciam axiomáticamente verdadeiras. Existe uma realidade objetiva "lá fora" (externa) que se faz valer independentemente de estarmos interessados nela; essa realidade funciona de acordo com determinadas leis imutáveis. É responsabilidade da ciência descrever tal realidade e revelar essas leis. Feito isso, a ciência pode ser usada para investigar a natureza em bene-

fício da espécie humana. Nós nos tornamos capazes de prever e controlar à vontade. Cada ato de investigação nos aproxima mais da compreensão da realidade suprema; em algum momento, seremos capazes de convergir para isso. No entanto, para compreender totalmente a realidade e suas leis, o investigador deve estar apto a manter-se distante (a uma distância neutra) do fenômeno sob investigação, para não o influenciar (o que nos impediria de ver as coisas tal como elas realmente são e funcionam) nem deixar que isso influencie seu juízo de valor. Devemos também ter consciência de que a natureza é ardilosa e tem várias táticas para obscurecer a busca da verdade. A confusão ruidosa e desastrosa da natureza em estado puro deve ser cuidadosamente controlada para evitar a confusão dos resultados que de outra forma com certeza ocorreria. Portanto, o investigador deve controlar o fenômeno, tanto por manipulação, como no laboratório, quanto estatisticamente, como é o caso na maioria dos estudos com seres humanos. Render-se ao controle é o mesmo que assegurar resultados falsos.

Praticamente todos os modelos de avaliação da primeira, segunda e terceira geração usam o paradigma científico para orientar seu trabalho metodológico (uma exceção é o Modelo de Crítica Artística, de Eisner, que alega seguir o paradigma humanista). Entretanto, essa extrema dependência em relação aos métodos da ciência produziu resultados lastimáveis. Primeiro, deu lugar ao que se chamou de "descontextualização", isto é, avaliar o objeto de avaliação como se ele não existisse em um dado contexto, mas apenas sob as condições cuidadosamente controladas que passam a vigorar depois que um projeto é implementado. Essas condições são instituídas na expectativa de que fatores locais inaplicáveis possam ser desprezados e resultados mais generalizáveis possam ser obtidos (é lamentável que muitas avaliações sejam contratadas para descobrir as qualidades generalizáveis dos objetos de avaliação). Posteriormente, defenderemos que a motivação para essa descontextualização é em todo caso equivocada, pois essas generalizações *não são possíveis*. Não obstante, por enquanto, vale enfatizar que, quando a atenção é direcionada apenas a fatores gerais, a situação local, tomada tal como em seu contexto original não descontextualizado, não pode ser bem atendida pelos resultados da avaliação. Além disso, os recursos necessários para instituir e manter os controles frustram outros usos para os quais a investigação possa ter sido posta a serviço; portanto, o leque de informações é truncado. Certamente, esse empenho

em deduzir verdades gerais valendo-se da descontextualização (controle) é um dos motivos pelos quais as avaliações são consideradas, na maioria das vezes, despropositadas em nível local, o que resulta na tão lamentável falta de uso dos resultados, falta sobre a qual nós, enquanto profissionais, aparentemente tanto gostamos de reclamar. Nenhuma das três gerações lida com esse problema.

Em segundo lugar, o comprometimento com o paradigma científico inevitavelmente parece gerar uma exagerada dependência em relação à mensuração quantitativa formal. O rigor que esse paradigma parece prometer apoia-se na “rigidez” dos dados que são fornecidos ao processo. Dados concretos pressupõem dados quantificáveis, dados que podem ser medidos com precisão e analisados com instrumentos matemáticos e estatísticos de grande eficácia. Os dados quantificáveis também atenuam o problema associado ao prognóstico e ao controle; eles podem ser facilmente inseridos em preceitos especialmente elaborados para essas atribuições. Depois de algum tempo, esses instrumentos de mensuração ganham vida própria; embora a princípio sejam concebidos para “operacionalizar” variáveis científicas, transformam-se, em última análise, nas próprias variáveis. Consequentemente, o que não pode ser mensurado não pode ser real. Nenhuma das três gerações lida com esse problema.

Em terceiro lugar, visto que os métodos da ciência nos prometem fornecer informações sobre o modo como as coisas realmente são, eles pretextam certa confiabilidade à qual é difícil resistir. Hannah Arendt (1963) falou sobre essa “coercitividade da verdade”. A verdade é negociável. No papel de avaliadores que utilizam o método científico, podemos assegurar a nossos clientes que a própria natureza nos forneceu os dados que nós, subsequentemente, apresentamos; ninguém discute com eles nem os contradiz. Nenhum de nossos valores, do cliente ou de qualquer outra pessoa, pode ser influenciado pelos resultados. Na função de avaliadores, podemos assumir a autoridade de mensageiros legítimos da qual a natureza nos vestiu. Se o *status quo* reflete as próprias leis da natureza, ele existe de certa forma por direito divino. Desse modo, é fácil ver como o método científico reforça e apoia as tendências gerencialistas que mencionamos anteriormente. Qualquer coisa que esteja sob avaliação e tenha o respaldo da avaliação positivista (científica) é pregada como a coisa certa a fazer. E nós, avaliadores, enquanto mensageiros, não somos responsáveis pelo que a natureza

decretou. Ambos, administrador e avaliador, tornam-se inexpugnáveis. Nenhuma das três primeiras gerações lida com esse problema.

Em quarto lugar, a utilização do método científico obsta formas alternativas de pensar sobre o objeto de avaliação. Visto que a ciência revela a verdade sobre as coisas, qualquer outra alternativa só pode estar errada. Avaliadores, clientes e do mesmo modo pessoas ou grupos de interesse são todos forçados a ser "fiéis verdadeiros/sectários", porque a ciência tem essa autoridade associada à sua capacidade de descobrir como as coisas realmente são e realmente funcionam. A pior coisa que se pode dizer em nossa cultura com relação a qualquer afirmação é que não existe nenhuma prova científica para apoiá-la; do contrário, quando *existe* prova científica, devemos aceitá-la acriticamente. Alternativas perfeitamente razoáveis não podem ser cogitadas de boa-fé. Nenhuma das três primeiras gerações lida com esse problema.

Por último, pelo fato de a ciência ser presumidamente isenta de valores, a adesão ao paradigma científico libera o avaliador de qualquer responsabilidade moral por suas ações. Ninguém pode ser acusado por estar apenas dizendo a verdade, por apresentar os fatos, por "descrevê-los tal como eles são vistos" ou por "deixar acontecer o que tiver de acontecer". É fácil argumentar que o avaliador não consegue controlar o modo como os resultados da avaliação são usados. É também fácil afirmar que o avaliador não tem nenhuma responsabilidade de fazer o acompanhamento a uma avaliação; sua função termina no momento em que apresenta seu relatório. Em todo caso, o avaliador (mensageiro) não pode ser responsabilizado por resultados (mensagem) que simplesmente reflitam o que existe na natureza. Nenhuma das três gerações responsabiliza o avaliador moralmente por qualquer resultado que provenha da avaliação ou pelo uso que se possa fazer dos resultados.

Uma abordagem alternativa

Esperamos ter deixado claro que uma abordagem alternativa à avaliação — aliás, no sentido exato do termo — é excessivamente necessária. Devemos propor uma abordagem, que designamos pelo título desajeitado, mas ainda assim bastante descritivo, de *avaliação construtivista responsiva* ou res-

pondente. O termo *responsivo* ou *respondente* é usado para designar uma forma diferente de *focalizar* uma avaliação, isto é, de escolher o que temos chamado de parâmetros e limites da avaliação. Nos modelos compreendidos pelas três primeiras gerações, os parâmetros e limites foram estabelecidos *a priori*; sua especificação, normalmente realizada por meio de negociações entre cliente e avaliador, faz parte do processo de concepção. Robert Stake cunhou o termo *avaliação preordenada* como forma de indicar essa qualidade *a priori*. A avaliação *responsiva*, também proposta pela primeira vez por Stake (1975), estabelece parâmetros e limites por meio de um processo interativo e negociado que envolve grupos de interesse e consome uma porção considerável do tempo e dos recursos disponíveis. É por esse motivo, dentre outros, que se diz que a concepção de uma avaliação *responsiva* é emergente.

O termo *construtivista* é usado para designar a metodologia de fato empregada para conduzir uma avaliação. Suas raízes encontram-se no paradigma de investigação, que é uma alternativa ao paradigma científico, e optamos por chamá-la de *construtivista*, mas existem vários outros nomes, como *avaliação interpretativa* ou *hermenêutica*. Todos esses termos oferecem uma percepção específica da natureza desse paradigma, e usaremos todos eles em diferentes partes deste livro. Nesta seção, apresentamos algumas das principais ideias que caracterizam, respectivamente, o método *responsivo de focalizar* e o método *construtivista de conduzir*. Cada um deles é investigado mais detalhadamente no desenrolar deste livro.

O método responsivo de focalizar. O algoritmo para qualquer processo de avaliação começa com um método para determinar que questões devem ser levantadas e quais informações devem ser coletadas. No caso da avaliação de primeira geração, identificam-se determinadas variáveis e as informações a serem coletadas consistem em pontuações específicas de instrumentos que presumivelmente medem essas variáveis (com frequência, nos ambientes escolares, elas se referem às notas de aproveitamento). No caso da avaliação de segunda geração, identificam-se determinados objetivos; as informações a serem coletadas consistem na avaliação da congruência entre o desempenho do aluno e os objetivos definidos. No caso da avaliação de terceira geração, modelos diferentes requerem informações diferentes; portanto, os modelos orientados à tomada de decisões, como o CIPP, exigem informações que atendam às decisões a serem toma-

das em tempo hábil (em geral isso requer que se colem informações comparáveis para cada uma das alternativas decisórias); os modelos sem referência a objetivos exigem informações sobre os “efeitos” experimentais; o modelo de crítica artística requer juízos de valor em relação a determinadas sinalizações críticas internalizadas pelos especialistas (*connoisseur*)/críticos por meio de sua formação e experiência; e assim por diante. Esses elementos de focalização — variáveis, objetivos, decisões etc. — podem ser chamados de “organizadores avançados”; o organizador que um avaliador está usando torna-se aparente tão logo ele levante algumas questões — por exemplo, “quais são seus objetivos” ou “a quais decisões esta avaliação deve fornecer informações”, e outras.

A avaliação responsiva também tem um organizador avançado: *reivindicações, preocupações e questões* com relação ao objeto de avaliação que são identificadas pelos interessados, isto é, pessoas ou grupos que a avaliação de certa forma coloca em risco. *Reivindicação* é qualquer alegação apresentada por um interessado (*stakeholder*) que seja favorável ao objeto de avaliação — por exemplo, que um método particular de instrução de leitura produzirá um ganho superior ao ganho de um ano nas notas de leitura obtidas em provas padronizadas em cada ano de utilização da sala de aula ou que um método particular usado pela polícia para lidar com telefonemas relativos a conflitos domésticos reduzirá substancialmente a reincidência dos infratores. *Preocupação* é qualquer afirmação que um interessado venha a apresentar que seja *desfavorável* ao objeto de avaliação — por exemplo, que as instruções sobre a utilização de computadores reduzem em grande medida a capacidade do aluno de realizar cálculos manualmente ou que o uso que o objeto de avaliação (avaliando) faz produzirá uma quantidade bem maior de “dever de casa” para os professores. *Questão* diz respeito a qualquer situação sobre a qual pessoas sensatas possam discordar — por exemplo, adoção do ensino sobre aids nas escolas elementares ou utilização das instalações escolares para ministrar aulas de religião. Diferentes grupos de interesse ou interessados acolherão diferentes reivindicações, preocupações e questões; é responsabilidade do avaliador deslindá-las e abordá-las na avaliação.

Sempre haverá vários e diferentes grupos de interesse. Em *Effective evaluation* [Avaliação eficaz] (Guba e Lincoln, 1981), identificamos três classes abrangentes, cada uma delas com alguns subtipos:

(1) Os *representantes* (ou agentes), as pessoas envolvidas na produção, utilização e implementação do objeto de avaliação — esses representantes abrangem:

- a. os desenvolvedores do objeto de avaliação;
- b. os patrocinadores locais, regionais e nacionais;
- c. os avaliadores das necessidades locais que identificam a necessidade que presumivelmente o objeto de avaliação melhorará ou eliminará;
- d. os tomadores de decisões que resolveram utilizar ou desenvolver o objeto de avaliação em nível local;
- e. os fornecedores de instalações, suprimentos e instrumentos;
- f. o cliente da avaliação em si (o contratante);
- g. o pessoal envolvido na implementação do objeto de avaliação, como os professores em sala de aula, funcionários de centros de reabilitação, oficiais de polícia, enfermeiras e outros.

(2) Os *beneficiários*, as pessoas que de alguma maneira tiram proveito do uso do objeto de avaliação — esses beneficiários abrangem:

- a. os beneficiários diretos, o “grupo-alvo”, as pessoas para as quais o objeto de avaliação foi criado;
- b. os beneficiários indiretos, pessoas cuja relação com os beneficiários diretos é intermediada, facilitada, aprimorada ou então influenciada positivamente;
- c. as pessoas que ganham com o fato de o objeto de avaliação estar sendo utilizado, como os indivíduos que publicam os conteúdos, os fornecedores contratados que oferecem os serviços necessários e outros.

(3) As *vítimas*, as pessoas que são afetadas negativamente pelo uso do objeto de avaliação (que podem abranger, em virtude de alguma falha no objeto de avaliação, um ou mais grupos considerados beneficiários).

Essas vítimas abrangem:

- a. os grupos excluídos sistematicamente do uso do objeto de avaliação, como jovens “normais” excluídos de programas especiais para bem-dotados;
- b. os grupos que sofrem efeitos colaterais negativos, como os alunos, e os pais, que são transportados de ônibus para uma escola distante para que os jovens desamparados possam ocupar seus lugares na escola original.

Avaliação de ...
c. as pessoas que são politicamente desfavorecidas pelo uso do objeto de avaliação, como aquelas que perdem poder, influência ou prestígio;

d. as pessoas que arcam com custos de oportunidade, com relação a oportunidades que foram renunciadas em consequência do uso atribuído ao objeto de avaliação, como aquelas que optariam por empregar os recursos necessários em algum outro empreendimento ou publicadores de conteúdos concorrentes.

A avaliação é responsiva não apenas pelo fato de procurar identificar diferentes visões entre os interessados, mas também porque responde às questões pertinentes à coleta de informações subsequente. É bem provável que diferentes interessados ou grupos de interesse defendam construções bastante distintas com respeito a qualquer reivindicação, preocupação ou questão em particular. Como veremos, uma das principais atribuições do avaliador é conduzir a avaliação de modo que cada grupo confronte e lide com as construções de todos os outros, um processo que chamaremos de hermenêutico-dialético. Nesse processo, algumas ou talvez várias reivindicações, preocupações e questões originais podem ser resolvidas sem recorrer a *novas* informações, isto é, informações que ainda não podem ser obtidas de um ou mais dos grupos interessados propriamente ditos. Visto que todo grupo lida com as construções apresentadas pelos demais, as construções de cada grupo alteram-se porque se tornaram mais bem fundamentadas e mais esclarecidas. Em teoria, nesse ponto do processo a avaliação responsiva procura obter consenso sobre todas as reivindicações, preocupações e questões, mas isso é raramente possível, quando muito. Restarão conflitos cuja resolução depende da introdução de informações externas, que caberá ao avaliador obter. Quando o avaliador tiver essas informações em mãos (tanto quanto puder obter), ele preparará uma *agenda de negociação*, tomando a frente para configurar uma sessão de negociação e atuar como moderador. Representantes de todos os interessados ou grupos de interesse pertinentes juntam-se ao avaliador, em um esforço conjunto, para tentar resolver o que está em pauta. As conclusões e recomendações finais provenientes dessa negociação (bem como aquelas alcançadas anteriormente no processo hermenêutico-dialético) são então alcançadas em *conjunto*; elas nunca são responsabilidade única e exclusiva do avaliador ou do cliente. As questões da agenda que não puderam ser

resolvidas mantêm-se como pontos de discórdia, obviamente, mas ao menos cada um dos interessados ou dos grupos de interesse perceberá qual é o conflito e que postura outros grupos assumem quanto a isso. O terreno está preparado para reciclar a avaliação. Essa iteração e reiteração são características da avaliação responsiva; as avaliações nunca se completam. Elas ficam em suspenso por motivos logísticos, como o momento oportuno para a tomada de uma decisão imperativa ou porque os recursos esgotaram-se.

A avaliação responsiva tem quatro fases, que podem ser reiteradas e sobrepor-se. Na primeira fase, os grupos de interesse ou interessados são identificados e solicitados a apresentar as reivindicações, preocupações e questões que desejarem. Na segunda fase, as reivindicações, preocupações e questões levantadas por cada grupo de interesse são apresentadas para que todos os outros grupos comentem, refutem, concordem ou reajam de qualquer maneira que lhes apraz. Nessa fase, várias reivindicações, preocupações e questões são resolvidas. Na terceira fase, as reivindicações, preocupações e questões que *não* foram resolvidas cumprem a função de organizadores avançados para a coleta de informações por parte do avaliador. O método exato de coleta de informações depende de o ponto de discórdia ser uma reivindicação (as informações podem ser coletadas para confirmar a reivindicação, por exemplo), uma preocupação (podem ser coletadas informações para averiguar até que ponto essa preocupação em questão se justifica) ou uma questão (podem ser coletadas informações que apoiem ou refutem cada um dos lados — e é possível haver mais de dois lados). *Essas informações podem ser quantitativas ou qualitativas.* A avaliação responsiva não descarta os métodos quantitativos, como muitos acreditam equivocadamente, mas lida com quaisquer informações que *possam* dar respostas à reivindicação, preocupação ou questão ainda não resolvida. Na quarta fase, ocorre a negociação entre os grupos interessados, sob a orientação do avaliador e por meio das informações avaliatórias que foram coletadas, com a finalidade de alcançar um consenso sobre o ponto de discórdia. Nem todos esses pontos serão resolvidos; os pontos restantes passam a ser centrais na próxima avaliação a ser empreendida, quando houver tempo, recursos e interesse.

Metodologia construtivista. A metodologia construtivista é a abordagem que propomos para substituir o método científico que caracterizou prati-

camente todas as avaliações realizadas no século XIX apoiava em um sistema de crenças praticamente oposto ao da ciência; um tipo de sistema de crenças que normalmente é referido como *paradigma*. De acordo com Michael Quinn Patton (1978, p. 203),

Paradigma é uma visão de mundo, uma perspectiva geral, uma forma de decompor a complexidade do mundo real. Como tal, os paradigmas estão profundamente enraizados na socialização de adeptos e praticantes: os paradigmas lhes indicam o que é importante, legítimo e sensato. Os paradigmas são também normativos, pois indicam aos praticantes o que fazer sem a necessidade de realizar amplas considerações existenciais e epistemológicas. Contudo, é a fisionomia do paradigma que constitui tanto seu ponto forte quanto seu ponto fraco — ponto forte porque torna uma ação possível, ponto fraco porque o motivo genuíno da ação fica velado nas suposições indiscutíveis do paradigma.

Não é possível comprovar ou refutar um paradigma em sentido absoluto, como não é possível, digamos, comprovar a existência de uma deidade nem o valor de um sistema acusatório (*adversarial system*) em uso nos tribunais nem o sistema crítico que caracteriza a crítica literária. Entretanto, mostraremos que é possível levantar questionamentos sobre o paradigma positivista que caracterizou a ciência contemporânea, questões essas fundamentais para que se possa afirmar que o paradigma positivista precisa ser substituído.

Acreditamos que o paradigma construtivista seja apropriado para essa atribuição. Ele em nada se parece com a ciência, em particular no que tange às suas suposições básicas, que são praticamente contrárias às da ciência. Porque, *ontologicamente*, ele nega a existência de uma realidade objetiva, afirmando, em vez disso, que as realidades são construções sociais mentais e que existem em número tantas construções quanto existem indivíduos (embora, sem dúvida, muitas construções venham a ser compartilhadas). Defendemos que a própria ciência é uma dessas construções; podemos admitir isso livremente ao panteão de construções somente se não formos solicitados a aceitar a ciência como a construção *certa* ou *verdadeira*. E enfatizamos que, se as realidades são construções, então não é possível haver, exceto por imputação mental, leis naturais imutáveis que

governam as construções, como leis de causa e efeito. *Epistemologicamente*, o paradigma construtivista nega a possibilidade do dualismo sujeito-objeto, propondo, em vez disso, que os resultados de um estudo existem precisamente porque existe uma *interação* entre o observador e o observado que cria exatamente o que provém dessa investigação. *Metodologicamente*, e em consequência das suposições ontológicas e epistemológicas já feitas, o paradigma naturalista rejeita a abordagem dominante e manipulatória (experimental) que caracteriza a ciência e a substitui por um *processo hermenêutico-dialético* que aproveita ao máximo — e leva em conta — a interação observador-observado para criar uma realidade construída que seja, tanto quanto possível, fundamentada e esclarecida em um determinado momento.

O leitor não deve deixar de observar a *afinidade especial* entre um paradigma de investigação que propõe uma metodologia hermenêutico-dialética e um modelo de avaliação que depende exatamente de tal processo para comprovar sua alegação de responsividade. O enfoque responsivo evoca uma metodologia construtivista, e a metodologia construtivista condiz precisamente com as necessidades do processo de investigação da avaliação responsiva.

As consequências da utilização da metodologia construtivista são surpreendentemente diferentes daquelas que passamos a presumir da investigação científica. Demonstraremos que ambas são formas de investigação *disciplinada*, no sentido do termo proposto por Cronbach e Suppes (1969), e isso significa que em ambas as metodologias é possível submeter à inspeção e comprovação pública “tanto as matérias-primas que entram no debate quanto os processos lógicos por meio dos quais eles foram reduzidos e reorganizados para tornar as conclusões dignas de crédito”. Com base nesse sistema de referência, argumentamos a favor de algo bastante diferente, comparativamente ao que as suposições científicas prenunciariam. Por exemplo (Guba, 1987):

– “Verdade” é uma questão de consenso entre construtores informados e esclarecidos, não de correspondência com uma realidade objetiva.

– Os “fatos” não têm nenhum significado, exceto dentro de uma estrutura valorativa; portanto, não é possível avaliar “objetivamente” nenhuma proposição.

- "Causa" e "efeito" não existem, exceto por imputação; por isso, a responsabilização é uma questão relativa e envolve igualmente todas as partes (entes) interagentes.

- Só é possível compreender os fenômenos no contexto em que eles são estudados; os resultados com respeito a um contexto não podem ser generalizados para outro; tampouco os problemas e as respectivas soluções podem ser generalizados de um ambiente para outro.

- As intervenções não são estáveis; quando forem introduzidas em um determinado contexto, no mínimo serão influenciadas (alteradas) por esse contexto na mesma proporção que tenderão a influenciá-lo.

- A mudança não pode ser engendrada; é um processo não linear que requer que se introduzam novas informações, e maior refinamento/esclarecimento em sua utilização, nas construções dos seres humanos envolvidos.

- A avaliação produz dados nos quais os fatos e os valores estão inextricavelmente relacionados. A apreciação é uma parte importante do processo de avaliação e oferece fundamentação para os significados atribuídos.

- Responsabilização é característica de um conglomerado de modeladores mútuos e simultâneos. Nenhum deles, tampouco um subconjunto desses modeladores, pode ser escolhido exclusivamente para louvar ou culpar.

- Os avaliadores são parceiros subjetivos dos grupos de interesse na criação factual dos dados.

- Os avaliadores são orquestradores de um processo de negociação que tenta alcançar um consenso sobre construções mais fundamentadas e esclarecidas.

- Os dados da avaliação provenientes de uma investigação construtivista não têm nenhum prestígio nem legitimação; eles simplesmente representam outra construção a ser levada em conta no processo em direção ao consenso.

Afirmações como as anteriores — e outras que talvez tenhamos feito e faremos em outra parte deste livro — à primeira vista parecem tão descartadas que são rejeitadas impensadamente. Contudo, parece haver uma mudança vigorosa na direção proposta por esse paradigma — muitas vezes, e equivocadamente, denominado paradigma qualitativo. Embora talvez nem todos — ou mesmo muitos — concordem em que o paradigma

naturalista deva ser o paradigma predileto, é surpreendente observar a variedade de campos em que tem havido divergências quanto ao positivismo científico e propostas de redirecionamento³. Nesse contexto, nossa proposição de realinhar a avaliação com um paradigma diferente do científico não parece tão incomum.

Os dilemas na aceitação da avaliação construtivista responsiva

A aceitação do método construtivista responsivo de conceitualização e de condução da avaliação envolve perdas e ganhos — embora o que se conta como ganho ou perda dependa da perspectiva segundo a qual alguém tem oportunidade de se manifestar.

Seguramente, os proponentes de modelos de avaliação mais convencionais estão propensos a considerar a mudança para a avaliação construtivista responsiva desastrosa e sujeita a inúmeras perdas. Em primeiro lugar, há uma aceitação implícita de que não pode haver nenhuma certeza sobre o modo como as coisas ocorrem; não existe nenhuma verdade objetiva para a qual as investigações possam convergir. Não é possível descobrir como as coisas realmente são ou de que modo elas de fato funcionam.

3 A seguir alguns exemplos — os 11 primeiros são extraídos do campo da educação e os últimos 12 de outros campos das ciências sociais: James W. Garrison, "Some principles of postpositivistic philosophy", *Educational Researcher*, 15, 1986, pp. 12-8; Kathleen Rockhill, "Researching participation in adult education — The potential of the qualitative perspective", *Adult Education*, 33, 1982, pp. 3-19; Linda F. Ettinger, "Styles of on-site descriptive research — A taxonomy for art educators", *Studies in Art Education*, 28, 1987, pp. 79-95; Mary Lee Smith, "Naturalistic research", *The Personal and Guidance Journal*, 59, 1981, pp. 585-9; J. T. Dillon, "The problems/methods/solutions of curriculum inquiry", *Journal of Curriculum and Supervision*, 1, 1985, pp. 18-26; K. J. Kantor; D. R. Kirby e J. P. Goetz, "Research in context — Ethnographic studies in English education", *Research in the Teaching of English*, 15, 1981, pp. 293-309; Marcy P. Driscoll, "Alternative paradigms for research in instructional systems", *Journal of Instructional Development*, 7, 1984, pp. 2-5; Jerome C. Harste, "Reading research — Portrait of a new paradigm", in R. J. Spiro (ed.), *Reading research in the 90's*, Hillsdale, Nova Jersey, Erlbaum, 1985; Barbara S. Spector, "Qualitative research — Data analysis framework generating grounded theory applicable to the crisis in science education", *Journal of Research in Science Teaching*, 21, 1984, pp. 459-67; Susan Stainback e William Stainback, "Broadening the research perspective in special education", *Exceptional Children*, 9, 1984, pp. 400-9; Judy Jax, "Ethnography — An approach to using interpretive science in vocational education research", *Journal of Vocational Research*, 9, 1984, pp. 8-19; Daniel S. Cochran e Janet A. Dolan, "Qualitative research — An

Esse grau de ambiguidade é quase insuportável. Se as avaliações não podem deslindar a verdade, de que vale realizá-las?

A mudança para a avaliação construtivista responsiva também pressupõe uma renúncia ao controle sobre o processo, supondo que os grupos de interesse ou interessados desempenhem funções igualmente definitivas em todos os estágios com o avaliador e o cliente. Essa perda de controle tem consequências metodológicas e políticas. Por um lado, se os indivíduos que normalmente não são especialistas em questões metodológicas ganharem importância enquanto tomadores de decisões, o exercício de sua prerrogativa pode ameaçar seriamente a adequação técnica do estudo. Além disso, se for concedido a essas pessoas o poder de tomar decisões metodológicas, elas simultaneamente ganharão controle político; a metodologia torna-se objeto de uma disputa decisiva entre grupos politicamente dissidentes.

Em terceiro lugar, comprometer-se com a avaliação construtivista responsiva significa abandonar a suposição de que é possível encontrar intervenções (tratamentos, programas, instrumentos, estratégias e outras) na avaliação que se demonstrem em grande medida equitativas a qualquer tarefa que se pretenda realizar, como atenuar o alcoolismo, melhorar a atuação das pessoas com subdesempenho, diminuir o recidivismo ou qualquer coisa que o valha. Se não houvesse nenhum fundamento para con-

alternative to quantitative research in communication", *Journal of Business Communication*, 21, 1984, pp. 25-33; Michael J. Piore, "Qualitative research techniques in economics", *Research Techniques in Economics*, 24, 1979, pp. 560-8; David A. Hollinger, "T. S. Kuhn's theory of science and its implications for history", *American Historical Review*, 78, 1973, pp. 370-93; Janet C. Harris, "Hermeneutics, interpretive culture research, and the study of sports", *Quest*, 33, 1981, pp. 72-86; Robert Grover e Jack Glazier, "Implications for application of qualitative methods to library and information science research", *Library and Information Science Research*, 7, 1985, pp. 247-60; Mary E. Duffy, "Designing nursing research — The qualitative-quantitative debate", *Journal of Advanced Nursing*, 10, 1985, pp. 225-32; Gareth Morgan e Linda Smircich, "The case for qualitative research", *Academy of Management Review*, 5, 1980, pp. 491-500; Joan C. Rogers, "Order and disorder in medicine and occupational therapy", *American Journal of Occupational Therapy*, 36, 1982, pp. 29-35; Kenneth J. Gergen, "The social constructivist movement in modern psychology", *American Psychologist*, 40, 1985, pp. 266-75; Glenn O. Haworth, "Social work research, practice, and paradigms", *Social Service Review*, 58, set., 1984, pp. 343-57; Leon Bramson, "Subjectivity in social research", capítulo 7, in Leon Bramson (ed.), *The political context of sociology*, Princeton, Nova Jersey, Princeton University Press, 1961. Esta lista não é de forma alguma representativa nem exaustiva (uma avaliação bem mais completa de publicações relacionadas a esse paradigma pode ser encontrada em Lincoln, 1989), mas mostra que o descontentamento com o paradigma convencional de investigação, detectado em vários campos, é predominante, havendo a proposição de uma variedade de alternativas.

siderar a generalizabilidade das intervenções ou de conceber intervenções com base nas relações já bem consolidadas de causa e efeito, haveria poucos motivos para acreditar, tal como alegariam os convencionalistas, que a sociedade pode enfim lidar com os vários problemas que evidentemente a acossam. Aceitar as premissas básicas reforçadoras da avaliação construtivista responsiva significa praticamente abandonar a expectativa de que algum dia será possível encontrar soluções para os problemas sociais. Se tudo tiver de ser adaptado a costumes específicos em contextos, culturas e níveis econômicos específicos, dentre outras coisas, a sociedade em breve será dominada por essa missão inacreditavelmente vasta e difícil.

Contudo, respondem os proponentes da avaliação construtivista responsiva, todos esses medos — da perda de verdades absolutas nas quais possamos fixar nossa esperança, da ambiguidade intolerável, da perda do controle experimental e político, de nossa incapacidade de encontrar soluções amplamente proficuas aos nossos problemas prementes — são em si apenas *construções* às quais seus construtores estão aprisionados em virtude de sua rígida adesão a suposições que evidentemente sobreviveram a sua utilidade e credibilidade. É precisamente em virtude de nossa preocupação em encontrar soluções universais que não conseguimos enxergar uma maneira de conceber soluções com significado e utilidade locais. É precisamente em virtude de nossa preocupação em ter esse controle que deixamos de dar poder e autonomia às próprias pessoas às quais supostamente estamos tentando servir.

A substituição da certeza que parece estar enraizada na metodologia convencional pelo relativismo característico da avaliação construtivista responsiva não implica uma postura do tipo “vale tudo”. Na verdade, essa mudança atrai especial atenção à questão sobre como alguém pode comparar uma construção com outra para determinar qual deve ser escolhida. Convencionalmente, tal comparação é feita com base na construção que mais se aproxima da realidade. Porém, quando se contradiz a possibilidade de de uma realidade objetiva, a norma desaparece, e outras distinções mais sutis e complexas são preconizadas. O imperativo moral do avaliador construtivista responsivo é estar continuamente alerta às contestações — na verdade, é procurá-las — contra a construção prevalecente (não importa o quanto ela se apoie no consenso) e prontificar-se a refinar, mudar ou mesmo rejeitar o que se acredita naquele momento, em favor de algo mais

que, após análise, pareça mais razoável e apropriado àqueles que estão em melhor posição de fazer esse juízo de valor. No mínimo, esse comprometimento com a avaliação construtivista responsiva substitui a arrogância tão facilmente presumida pelo convencionalista — convencido de que descobriu algo sobre a “realidade” — pela humildade própria do discernimento de que *nunca* é possível saber como as coisas “realmente” são; de que a construção que se faz sobre como as coisas são na verdade é criada pela própria investigação e não é determinada por certa “natureza” misteriosa. Substituir a certeza pela relatividade, o controle pela concessão de poder (*empowerment*), a explicação generalizada pela compreensão local e a arrogância pela humildade parece ser uma série de benefícios evidentes para o avaliador da quarta geração.

Você, leitor, terá de decidir por si mesmo de que modo deseja computar os ganhos e as perdas.