

# Desenvolvimento de medidas em avaliação de treinamento<sup>1</sup>

**Jairo Eduardo Borges-Andrade**

Universidade de Brasília

## Resumo

O texto descreve um conjunto de esforços empreendidos para desenvolver medidas de avaliação de treinamento. Ele é organizado em torno de um modelo de cinco níveis de avaliação: “reação”, “aprendizagem”, “comportamento no cargo”, “organização” e “valor final”. São discutidos os desafios metodológicos de cada um desses níveis de avaliação e apresentados exemplos de como os problemas de mensuração têm sido resolvidos, com o desenvolvimento de instrumentos e a implementação de estratégias de coleta e análise de dados. Os desafios por serem enfrentados mais sistematicamente são apontados e são indicados os maiores avanços realizados.

Palavras-chave: Medidas em avaliação de treinamento, Satisfação com treinamento, Aprendizagem em treinamento, Impacto de treinamento no desempenho individual, Impacto de treinamento no desempenho organizacional.

## Abstract

*Development of measures in training evaluation.* The text describes a set of efforts that have been undertaken in order to develop training evaluation measures. It is organized around a five-assessment levels model: “reactions”, “learning”, “behavior at job”, “organization” and “final value”. Methodological challenges of each of these assessment levels are discussed. Instruments development and implemented data collection and analysis strategies are presented, as examples of how measurement problems are being solved. Challenges that have not yet been more systematically faced are pointed and the major accomplished advances are indicated.

Key words: Measurement in training evaluation, Training satisfaction, Learning in training, Training impact on individual performance, Training impact on organizational performance.

## Justificativas e conceitos

**E**xiste crescente demanda por avaliação, nas organizações que tradicionalmente fizeram altos investimentos em treinamento, bem como naquelas que mais recentemente descobriram os valores estratégicos do conhecimento e do esforço na contínua qualificação de seu pessoal. O aumento da produção científica e tecnológica na área, uma recente resposta a essa demanda, indica que não é mais possível continuar simplesmente realizando avaliações de treinamento e fazendo pesquisas sobre elas, ignorando-se os dilemas que são enfrentados cada vez que se elabora um instrumento de medida e se coleta e analisa dados.

São aqui apresentados e discutidos os instrumentos e procedimentos que um núcleo brasileiro<sup>2</sup> tem usado para fazer suas avaliações e pesquisas. Pretende-se relatar resumidamente, analisar e discutir experiências realizadas que

estão concretizando diferentes formas de apreensão dos processos e produtos do treinamento nas organizações. O prisma será o de contemplar a diversidade existente, buscando organizar as experiências relatadas com base num modelo de “níveis de avaliação” hegemônico na área. Serão destacadas as práticas em uso e apontados os seus problemas, de forma a contribuir com futuros estudos e experiências de avaliação, com o foco no método.

O leitor deve ser alertado, no entanto, para o fato de que o presente texto não pretende ser um manual para o desenvolvimento de medidas em avaliação de treinamento que substitua a consulta aos textos aqui revisados. Consultas diretas aos textos citados são altamente recomendáveis, uma vez que esta revisão faz apenas considerações e relata resultados gerais.

De acordo com Borges-Andrade e Abbad (1996), uma das características essenciais do conceito de “Treinamento”

(T) é a noção de que ele representa um esforço despendido pelas organizações para propiciar oportunidades de aprendizagem aos seus integrantes. Ele está tradicionalmente relacionado à identificação e superação de deficiências no desempenho de empregados, preparação de empregados para novas funções e adaptação da mão de obra à introdução de novas tecnologias no trabalho. O conceito de “Desenvolvimento” (D), na literatura clássica da área, é compreendido como mais abrangente, incluindo ações organizacionais que estimulam o livre crescimento pessoal de seus membros, que não visam necessariamente a melhoria de desempenhos atuais ou futuros (Nadler, 1984).

Devido às rápidas e vertiginosas mudanças tecnológicas, econômicas e sociais que caracterizam o atual mundo do trabalho e das organizações, T & D têm crescido de importância. Através deles, são adquiridas habilidades motoras ou intelectuais e informações e desenvolvem-se estratégias cognitivas e atitudes, que poderão tornar o indivíduo mais competente para desempenhar vários papéis, no presente ou no futuro e em diversas organizações. No caso de “Treinamento”, essa aquisição seria feita de uma forma mais sistemática, utilizando-se uma tecnologia instrucional, enquanto “Desenvolvimento” estaria mais baseado na autogestão da aprendizagem. Ao analisar os problemas conceituais da área, Bastos (1991) sugere que uma das maneiras de diferenciar T & D seria pelos critérios da intencionalidade em produzir melhorias de desempenho e do controle exercido pela organização sobre o processo de treinamento.

A noção de crescimento pessoal contínuo, inicialmente colocada como um assunto de decisão pessoal, tornou-se um requisito organizacional que dele retirou a noção de “escolha livre e pessoal”, embora tenha sido mantido ou fortalecido o discurso de que se trata de autogestão da aprendizagem. O desenvolvimento de competências pessoais diversificadas se transformou em estratégia organizacional, cuja efetivação leva evidentemente a maiores expectativas de controle e a uma redução drástica do âmbito do que pode ser realmente considerado como autogestão. Por outro lado, um treinamento cuidadosamente formulado, para resolver problemas específicos de desempenho, de um indivíduo ou uma categoria ocupacional, pode ser visto, por outros indivíduos ou categorias, como uma oportunidade de desenvolvimento que garantiria melhor empregabilidade no mercado interno ou externo à organização. As noções de sistematização, de tecnologia e de intencionalidade, que pareciam mais restritas ao conceito de “Treinamento”, podem ser sorvidas pelo conceito de “Desenvolvimento”, quando este é hoje posto em prática como estratégia organizacional e num contexto de grande competição por trabalho e emprego. Da mesma forma que as chamadas aprendizagens individual e organizacional precisam coexistir num mesmo espaço de

tempo e lugar, T & D parecem estar intimamente relacionados e seus conceitos podem estar perdendo as claras fronteiras que os separavam há menos de duas décadas.

T & D podem ser vistos como um **sistema**, integrado por três elementos: (a) avaliação de necessidades; (b) planejamento do treinamento e sua execução e (c) avaliação do treinamento. Do primeiro para o segundo e deste para o terceiro, os referidos elementos mantêm entre si um constante fluxo de informações e produtos, sendo que o subsistema “avaliação de treinamento” seria o principal responsável pelo provimento de informações que garante a retroalimentação e, portanto, o aperfeiçoamento constante do sistema. Esta avaliação pode ser definida como um processo que inclui sempre algum tipo de coleta de dados usados para se emitir um juízo de valor a respeito de um treinamento, ou um conjunto de treinamentos.

Hamblin (1978) propôs que uma avaliação de treinamento deveria seguir cinco níveis: (I) **reação**, que levanta atitudes e opiniões dos treinandos sobre os diversos aspectos do treinamento, ou sua satisfação com o mesmo; (II) **aprendizagem**, que verifica se ocorreram diferenças entre o que os treinandos sabiam antes e depois do treinamento, ou se os seus objetivos instrucionais foram alcançados; (III) **comportamento no cargo**, que leva em conta o desempenho dos indivíduos antes e depois do treinamento, ou se houve transferência para o trabalho efetivamente realizado; (IV) **organização**, que toma como critério de avaliação o funcionamento da organização, ou mudanças que nela possam ter ocorrido em decorrência do treinamento, e (V) **valor final**, que tem como foco a produção ou o serviço prestado pela organização, o que geralmente implica em comparar custos do treinamento com os seus benefícios.

Em qualquer planejamento de treinamento poderiam ser encontrados, explicitados ou não, objetivos esperados relativos a cada um desses níveis (ver parte esquerda da Figura 1). A lógica do planejamento organizacional seria, desse modo, a de formular os objetivos do nível de valor final e deles derivar os objetivos do nível seguinte, e assim por diante, até os objetivos do nível I. Para alcançá-los, existiriam muitos possíveis tipos de ações, sendo que a maioria delas não envolveria treinamento. Isso sugere, portanto, que T & D não deveriam ser considerados panacéias para resolver todos os problemas organizacionais.

A seqüência em que ocorreriam os efeitos seria iniciada pelas reações dos treinandos e terminaria no nível V (ver parte direita da Figura 1). O modelo sugere uma cadeia de relações de determinação, em que os resultados do primeiro nível seriam os maiores responsáveis pela aprendizagem dos indivíduos, que produziriam mudanças no seu desempenho no trabalho e assim por diante. Em cada um desses níveis, no entanto, se agregariam outras variáveis interferentes, de

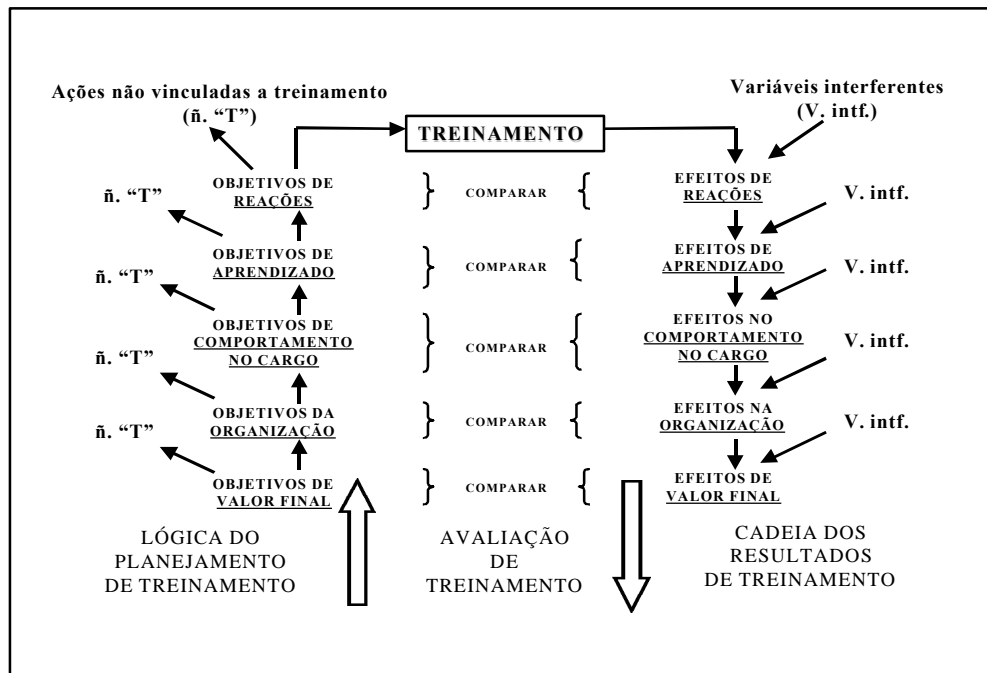


Figura 1. Níveis de avaliação de treinamento (adaptação de Hamblin, 1978)

modo que os efeitos atribuíveis ao treinamento se tornariam dependentes de um conjunto gradativamente mais amplo de fatores, o que dificultaria mais a sua detecção.

O papel da avaliação de treinamento seria o de coletar dados relativos aos efeitos nos diferentes níveis e compará-los com aqueles que seriam esperados (ver parte central da Figura 1). Contudo, dados nem sempre seriam colhidos em todos os níveis.

### Medidas de reação

A construção de medidas de reação reclama a resposta a três questões: O que medir? Com o que medir? Como medir?

A resposta sobre o que medir passa por uma outra decisão, pois se pode obter uma medida global, com um único item, ou múltiplas medidas de reação, que direcionam o foco do respondente para distintos aspectos do treinamento. O uso de um item, como indicador global, apresenta riscos quanto à sua confiabilidade. O uso de vários itens pode ser muito facilitado por um quadro de referências que guie a elaboração de questões que cubram apropriadamente todos os aspectos do treinamento a serem considerados. O Modelo de Avaliação Integrado e Somativo - MAIS, representado pela Figura 2, tem servido bem a esse propósito, pois sua base teórica privilegia justamente os aspectos instrucionais e administrativos que geralmente produzem reações nos treinandos. Ele sugere que uma avaliação deve considerar múltiplas variáveis, classificadas em cinco componentes:

insumos, procedimentos, processos, resultados e ambiente (com quatro subcomponentes). Componentes e subcomponentes que são vizinhos (separados por linhas contínuas ou pontilhadas), principalmente no sentido esquerdo - direito, seriam aqueles em que haveria maior probabilidade de existirem relações de dependência.

**Insumos**, no referido modelo, são definidos como os fatores físicos e sociais e estados comportamentais, geralmente associados ao treinando, anteriores ao treinamento e que podem afetar sua realização (ex.: experiências anteriores na organização e motivação antes do treinamento). **Procedimentos** são as operações realizadas para produzir os resultados instrucionais, geralmente controladas pelo instrutor ou por algum meio de entrega da instrução (ex.: seqüência de objetivos, exercícios realizados, retroalimentação recebida pelo treinando). **Processos** são definidos como as ocorrências resultantes da realização dos procedimentos e geralmente associadas a desempenhos intermediários dos treinandos, mas podendo prever resultados finais (ex.: relações interpessoais desenvolvidas no treinamento). **Resultados**, outro componente do MAIS, compreendem o que foi aprendido pelos treinandos ou por eles alcançado ao final do treinamento (ex.: qualquer habilidade adquirida ou atitude desenvolvida). Ele corresponderia ao nível II do modelo de Hamblin (1978).

O modelo MAIS se encerra com o componente **ambiente**, que é dividido em quatro subcomponentes: **necessidades** existentes no ambiente de trabalho e que geralmente são

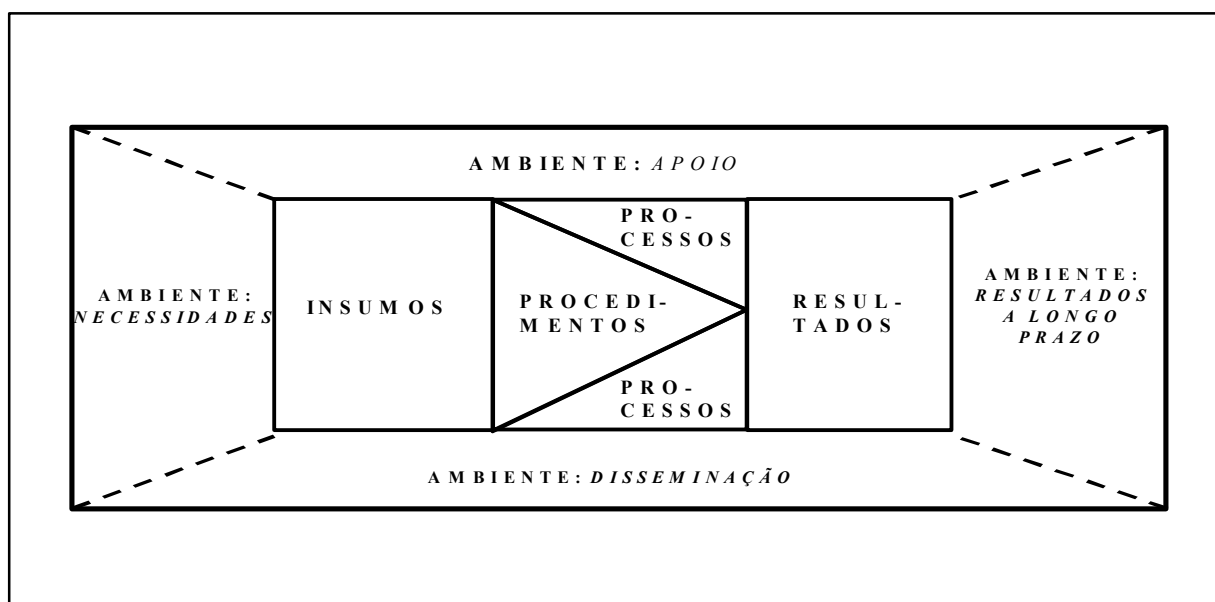


Figura 2. Modelo de Avaliação Integrado e Somativo (Borges-Andrade, 1982)

definidas como lacunas ou discrepâncias entre desempenhos esperados e apresentados, podendo influenciar os insumos; **apoio**, que abrange variáveis no lar, na organização ou na comunidade, que influenciam diretamente os insumos, procedimentos e processos, e indiretamente a aprendizagem e os resultados a longo prazo (e.g., instalações, proporção treinando/instrutor, suporte social para transferência do aprendido ao trabalho); **disseminação**, que arrola outros aspectos que podem contribuir para a procura ou escolha do treinamento, tais como os meios e estratégias usados para sua divulgação e **resultados a longo prazo**, ou as conseqüências ambientais do treinamento medidas após algum tempo, estando neles inseridos os níveis III a V propostos por Hamblin (1978).

A construção dos itens, para mensuração das reações, é feita com base no quadro de referências oferecido por esse modelo e levando em consideração as peculiaridades do contexto organizacional e do sistema de “T & D” em funcionamento. Seria ideal possuir instrumentos contendo itens de reação aplicáveis a qualquer contexto organizacional, mas isso teria o alto custo da perda da especificidade dos mesmos. Assim, a tendência tem sido a de construir questionários específicos para cada organização, embora tomando o MAIS como modelo ou quadro de referências. Um exemplo disso foi a experiência desenvolvida por Lima e Borges-Andrade (1985), que construíram um questionário para avaliar reações a uma grande variedade de treinamentos oferecidos pela Embrapa. Contudo, esses autores não investigaram a validade de construto do seu instrumento, nem a confiabilidade dos escores que poderiam ser obtidos a partir dos itens elaborados com base no modelo MAIS. Fazer isso

para cada organização é um desafio. É preciso contar com treinamentos variados e em grande quantidade, uma organização disposta a se submeter a esse processo e uma equipe bem capacitada. Experiências bem documentadas e com indicadores apropriados de validade de construto (usando análises fatoriais) e de confiabilidade (calculando alfas de Cronbach) foram feitas na Telebrás, por Alves, Pasquali e Pereira (1999), e no Tribunal de Contas da União (TCU), por Abbad, Gama e Borges-Andrade (2000).

Essas experiências demonstraram que é possível desenvolver instrumentos de reações confiáveis e extrair fatores que mantêm uma correspondência razoável com o quadro de referências que os inspirou e que se ajustam às peculiaridades das organizações e de seus sistemas de “T & D”. Contudo, essa última vantagem é também um ponto fraco, pois em todos os casos não se pode garantir maior generalidade dos instrumentos, ou que eles se comportariam de maneira equivalente em outros contextos. Outro problema tem a ver com o próprio MAIS, que pretende abranger todos os aspectos do treinamento, tornando-se um recurso que produz “explosões de criatividade” nos elaboradores de questões e resultando em questionários inicialmente muito extensos. Eles provocam resistências mesmo nas aplicações de teste dos mesmos, até que as análises dos primeiros dados obtidos possam sugerir o que eliminar.

Há duas outras questões a serem consideradas. Com o que medir? As respostas podem ser com lápis e papel, entrevistas ou observações. Como medir? Isso pode ser feito com registros de observação. Ou com questões fechadas ou abertas. As experiências têm sido com lápis e papel, o que exige um certo nível de escolaridade dos respondentes, e

quase sempre questões fechadas, o que limita as chances de serem detectadas ocorrências não previstas, mas reduz o tempo de análise e aumenta a confiabilidade dos resultados. Os questionários geralmente incluem uma questão aberta, que exige uma análise de conteúdo trabalhosa, utilizando-se um conjunto de categorias com base no próprio MAIS. As análises quantitativas e qualitativas têm usualmente um alto grau de correspondência, já que as várias questões fechadas podem influenciar o que é escrito nas questões abertas. Isto é, na verdade, um outro problema de mensuração.

Para evitar os problemas dos questionários de lápis e papel, Abbad, Sallorenzo e Gama (1998) construíram e aplicaram um roteiro de observação do instrutor em sala de treinamento. Os resultados finais obtidos foram animadores, demonstrando que é possível a obtenção de índices de concordância entre observadores acima de 90%. Mas o método exige um grande esforço de coordenação e treinamento de observadores e é limitado aos aspectos diretamente observáveis, não sendo sensível para a detecção da satisfação e opiniões dos participantes. Ao invés de focalizarem o desempenho do instrutor, Pantoja, Lima e Borges-Andrade (1999) relatam um método de análise de documentos relativos a programas de cursos, utilizando uma lista de verificação, abrangendo aspectos relacionados ao planejamento instrucional. Nesse caso, o problema foi que as naturezas do instrumento e do procedimento de coleta de dados inviabilizaram um tratamento estatístico que verificasse a confiabilidade dos dados obtidos e a validade de construto das medidas. O método compartilha, com o anterior, a desvantagem de que não revela a satisfação dos participantes.

Em resumo, para responder às três questões colocadas no início desta seção, é necessário levar em consideração a natureza da organização e do seu sistema de "T & D", os tipos de clientela desse sistema e os recursos disponíveis. As experiências já realizadas demonstram que se avançou no sentido de enfrentar os desafios da construção de medidas de reações, embora as soluções desenvolvidas tenham igualmente produzido outros desafios. Há dilemas que persistirão, pois as alternativas de resposta tendem a atender a uma ou duas dessas considerações, mas deixam as demais sem cobertura.

### *Medidas de aprendizagem*

O desenvolvimento de medidas de aprendizagem tem como primeiro desafio a seguinte questão: os parâmetros, contra os quais será comparado o desempenho de cada treinando, devem ser fornecidos pelo desempenho do grupo de treinandos ou pelos desempenhos explicitados nos objetivos do treinamento? No primeiro caso, a construção dos itens é geralmente feita a partir de uma amostragem dos conteúdos do programa, enquanto no segundo, o item é deduzido dire-

tamente dos objetivos formulados. O primeiro vem de uma tradição de seleção: procuram-se os mais aptos. O segundo vem de uma tradição de ensino: deseja-se determinar quem aprendeu. Embora a resposta pareça óbvia, pois a finalidade imediata de um treinamento é a aprendizagem, o fato de que muitos treinamentos não possuem objetivos formulados em termos de desempenho impede que a segunda tradição seja seguida. Então, elaboram-se um conjunto de itens baseado numa amostragem do conteúdo incluído no programa. Assim, o controle do processo dificilmente fica nas mãos do avaliador, passando para as do especialista em conteúdo, o que impede uma apropriada construção e análise de itens. Em muitas organizações, é tal a autonomia dada ao instrutor, que este ou os responsáveis pelo sistema de "T & D" considerariam inconveniência, ou intromissão, a participação do avaliador ou pesquisador na elaboração de itens.

Outro problema é que uma boa medida de aprendizagem deveria comparar desempenhos antes e depois do treinamento. Isto é dificilmente viável, pois quase sempre não existem baterias de itens efetivamente similares para serem usadas nestes dois momentos. Além disso, começar um treinamento com uma avaliação pode ser fator de constrangimento em muitas culturas organizacionais. Mesmo a avaliação realizada somente ao final do treinamento é vista como imprópria ou considerada como algo inerente ao ensino formal e não a organizações de trabalho. Algumas estratégias têm sido utilizadas para a obtenção de quaisquer indicadores de aprendizagem. Pantoja et al. (1999) e Abbad (1999) utilizaram resultados de pós-testes construídos e corrigidos pelos próprios instrutores e encontraram poucas correlações significativas ou correlações de Spearman-Brown abaixo de 0,30 entre esses indicadores e variáveis relativas a outros componentes do MAIS ou a outros níveis do modelo proposto por Hamblin (1978). Em vez de questionar as hipóteses desses modelos, os autores preferiram reconhecer a fragilidade de suas medidas de aprendizagem, pois não tiveram acesso aos processos de construção e análise dos itens, nem controle sobre os mesmos. Não tendo sequer conseguido essas medidas, mas podendo contar com objetivos de treinamento formulados em termos de desempenho, Borges-Andrade, Morandini e Machado (1999c) e Lima e Borges-Andrade (1985) recorreram a auto-relatos de aprendizagem, em que listavam os objetivos de ensino e solicitavam aos respondentes que julgassem sua aquisição durante o treinamento. A não ser pelo fato de terem verificado consistência interna entre esses julgamentos, seus resultados não foram mais animadores, pois foram auto-avaliações "pós-fato".

São grandes as barreiras que impedem o desenvolvimento de instrumentos de mensuração no nível de aprendizagem. O que foi feito até agora permite concluir que os avanços absolutos foram pequenos. Comparando-se o que

foi alcançado neste nível com aquilo já relatado no nível de reação, conclui-se que também não houve avanços em termos relativos. Quando o balanço das perdas e ganhos é feito, chega-se a acreditar que foram produzidas mais zonas de sombra do que de luz. Há que considerar que as pesquisas sobre avaliação de treinamento tiveram como focos de interesse o nível anterior (reação) ou o próximo (desempenho no cargo) e os indicadores do nível de aprendizagem foram coadjuvantes ou complementares. Para que progressos substantivos possam ocorrer neste nível, um esforço mais metódico e uma visão finalista são necessários.

### *Medidas de comportamento no cargo*

Tal como na mensuração de reações, no processo de desenvolvimento de medidas no nível de comportamento no cargo já existe sistematização para a tomada de decisões. As seguintes questões devem ser respondidas: (1) O que medir? - Comportamento ou resultado? (2) Em que nível de complexidade? - De “profundidade” ou “largura”? (3) Como medir? - Observar ou perguntar? (4) Quem deve fornecer os dados? - Treinandos, supervisores, colegas ou clientes? As respostas, como demonstrado a seguir, dependem de condicionantes como disponibilidade de recursos humanos e financeiros no subsistema de avaliação, natureza dos objetivos de treinamento, tipo de cargo ou função, cultura da organização e tipo de clientela a ser avaliada.

A decisão sobre “o que medir” é tomada após uma análise dos objetivos formulados para o treinamento. Na maioria dos casos, essa análise sugere que o foco permaneça em comportamentos, mas há também situações em que ele fica em resultados deduzidos a partir dos comportamentos descritos nos objetivos. Estes, quando existentes, são freqüentemente escritos num nível de especificidade apropriado para planejamento da instrução, mas excessivamente detalhado para uma avaliação no cargo. Uma eliminação dos detalhes ou dos objetivos estritamente instrucionais é necessária, sendo que as justificativas utilizadas para implementar o treinamento ou o diagnóstico de necessidades realizado podem ajudar muito nessa decisão.

Obtida uma lista de itens que descrevem comportamentos ou resultados do treinamento esperados no trabalho, tem sido solicitado aos respondentes que opinem, com o apoio de uma escala tipo *Likert*. Borges-Andrade e Siri (2000) utilizaram uma escala de julgamento de nível de impacto no trabalho, avaliando um programa de treinamento oferecido para membros de organizações operando na América Latina e Caribe, que tinha uma lista bem clara de comportamentos e resultados esperados, na área de planejamento, acompanhamento e avaliação da pesquisa agrícola. Borges-Andrade, Azevedo, Pereira, Rocha e Puente (1999a), no Banco do Brasil, e Pilati, Borges-Andrade e Azevedo (1999),

numa empresa privada de televisão por assinatura, aplicaram uma escala de freqüência de utilização, no trabalho, do aprendido em vários treinamentos oferecidos em todo o Brasil. A escala foi sempre a mesma, mas foram muitas as listas de comportamentos esperados no cargo, já que os treinamentos variaram muito quanto à natureza de seu conteúdo (técnicos, operacionais e gerenciais) e quanto à natureza das capacidades ensinadas (habilidades intelectuais, informações verbais e atitudes).

As tentativas realizadas por esses autores, no sentido de verificar a validade de construto e a confiabilidade dos instrumentos, resultaram em estruturas unifatoriais e alfas de Cronbach entre 0,80 e 0,90. Contudo, quando o número de treinandos por treinamento é pequeno, esse tipo de análise estatística fica inviabilizado, pois ele precisa ser feito para cada lista de comportamentos esperados. Isso só pode ser superado quando são realizadas avaliações de poucos treinamentos oferecidos em caráter massificado ou em organizações de grande porte.

O primeiro grande desafio ocorre quando não existem objetivos formulados em termos de comportamentos esperados do treinando. Isto acontece porque o planejamento segue uma abordagem que rejeita essa formulação de objetivos, ou só os formula em termos do que é esperado dos instrutores, ou só os formula no decorrer do próprio treinamento, após discussão com os participantes, ou só faz listagens de conteúdos a serem abordados. Outras razões, bem mais prosaicas, são as de que o sistema de “T & D” é tão desorganizado que não existe documentação registrando os planejamentos ou é totalmente terceirizado, a ponto de não fazer sentido uma política de arquivamento desse tipo de informação. Várias experiências foram desenvolvidas para tentar resolver esse problema.

A primeira foi a de construir perguntas que servissem para avaliar o impacto de qualquer treinamento no trabalho, independentemente da abordagem de “T & D” ou da política adotada pelas organizações. Lima, Borges-Andrade e Vieira (1989) identificaram três importantes indicadores e solicitaram aos respondentes que utilizassem uma escala de concordância tipo “Likert”, para julgar o impacto do treinamento que tinham realizado. A mesma estratégia foi depois usada por Leitão (1994) e Paula (1992). As medidas pareciam promissoras, pois contornavam o problema apresentado, podendo ser usadas em qualquer organização e para qualquer treinamento, havendo evidências de validade de conteúdo. Mas careciam de uma melhoria substancial, verificando-se, por exemplo, sua validade de construto e confiabilidade.

Abbad (1999) envidou esse esforço de melhoria, tomando os itens já utilizados nessas experiências e a eles acrescentando outros que emergiram de um levantamento exploratório feito com entrevistas. As análises estatísticas

realizadas revelaram uma estrutura unifatorial e índice de confiabilidade (alfa de Cronbach) acima de 0,90. Outras tentativas, já usando o novo instrumento desenvolvido, foram descritas por Borges-Andrade, Gama e Oliveira-Simões (1999b), Borges-Andrade et al. (1999c), Pantoja et al. (1999) e Martins, Pinto Jr. e Borges-Andrade (1999), que encontraram resultados de análises fatoriais e de confiabilidade muito similares aos da autora, o que possibilitou concluir sobre a generalidade dos mesmos para outras organizações.

Outro esforço foi realizado por Alves e Tamayo (1993), que não tomaram como base os indicadores anteriores. Eles realizaram toda a construção de seu instrumento a partir da opinião de treinandos e gerentes da Telebrás e encontraram estruturas multifatoriais, incluindo dimensões tais como motivação, relacionamento, autovalorização, atitude crítica e organização do trabalho, a maioria delas com índices de confiabilidade (alfas de Cronbach) acima de 0,80. Não existe registro, no entanto, de que esse instrumento tenha sido posteriormente aplicado em outras organizações.

Essas experiências de construção de instrumentos com itens idênticos parecem ser apropriadas para situações em que a natureza dos objetivos (quando eles existem) não permite a elaboração de listas de desempenhos esperados para cada treinamento ou quando o sistema de “T & D” é extremamente grande e a organização deseja avaliar comparativamente todos os eventos realizados. Contudo, a tomada de decisão no sentido de elaborar tais instrumentos exige que se disponha de recursos financeiros, de pessoal capacitado para construí-los e de uma cultura organizacional que compreenda os procedimentos de pesquisa e esteja disposta a aguardar pelos seus produtos. Isso pode ser contornado com o uso do instrumento já desenvolvido, cuja generalidade foi demonstrada, mas é preciso alertar que ele não permite que se leve em conta medidas de comportamento no cargo específicas de cada ambiente organizacional.

Uma segunda alternativa para resolver o desafio antes colocado é fazer um levantamento exploratório, de caráter qualitativo, utilizando entrevistas e leitura de documentos e abrangendo todas as pessoas que estiveram envolvidas com o treinamento. Assim, é feito um grande esforço de recuperação e sistematização de informações que estão dispersas ou existentes somente na memória das pessoas. A organização que deseja fazer isso paga, em termos dos recursos financeiros e humanos que deverá utilizar no subsistema de avaliação, o preço de não ter documentado e organizado seu subsistema de planejamento de “T & D” ou de ter decidido terceirizá-lo. Mesmo assim, essa alternativa pode não servir para qualquer tipo de cargo ou função, clientela de treinamento ou cultura organizacional.

Um exemplo de implementação dessa alternativa foi descrito por Bastos, Fernandes e Viana (1999), que avalia-

ram um programa que pretendia a transferência para o trabalho de complexas e subjetivas dimensões de habilidades pessoais esperadas para lidar com demandas estressantes num hospital, que deveriam ser fortalecidas em um “grupo de crescimento”. O levantamento de seus indicadores foi feito através de entrevistas individuais, análises de dados secundários e do uso da técnica de grupo focal. Na parte quantitativa, as respostas foram dadas numa escala tipo “Likert”, que solicitava julgamentos sobre o nível de utilização, no trabalho, do que tinha sido aprendido no programa. Contudo, o elevado número de indicadores e a quantidade reduzida de respondentes impossibilitaram a verificação da validade de construto e da confiabilidade dos instrumentos.

Além dos indicadores poderem se diferenciar em termos de comportamentos e resultados, seu nível de complexidade pode ser de “profundidade” ou “largura” (Hamblin, 1978). No primeiro caso, o foco da avaliação é estritamente no que está previsto no programa do treinamento, mesmo que isso não esteja escrito e que seja necessário fazer um levantamento exploratório. Portanto, os instrumentos de coleta de dados limitam-se a questionar o uso dos conhecimentos e habilidades aprendidas ou atitudes desenvolvidas, ou o seu impacto no trabalho. Várias pesquisas já descritas aqui seguiram essa estratégia. Entretanto, sob certas culturas e políticas organizacionais é possível, ou é preciso, verificar se existem evidências de que o adquirido em treinamento está tendo impacto em dimensões do desempenho individual que vão além daquelas diretamente relacionadas àqueles conhecimentos, habilidades e atitudes previstos em programas de treinamento.

Borges-Andrade et al. (1999a) usaram o instrumento em “profundidade” já descrito e testado em termos de validades de conteúdo e de construto e desenvolveram um outro em “largura”. Para isso, identificaram descrições de desempenhos esperados de todos os empregados do Banco do Brasil, denominadas de fatores corporativos de desempenho, discutidos durante o planejamento estratégico da empresa e amplamente conhecidos como parâmetros de avaliação de desempenho individual. Essas descrições foram incorporadas no questionário e deveria ser feito um julgamento, utilizando-se uma escala tipo “Likert”, a respeito do nível de impacto do treinamento em cada um dos fatores listados.

Pilati et al. (1999) utilizaram uma estratégia de mensuração em “profundidade” e “largura” extremamente similar a esta experiência do Banco do Brasil, para avaliar treinamentos operacionais na já citada empresa privada de televisão por assinatura. Em ambas experiências, foram obtidos indicadores de confiabilidade (alfas de Cronbach) acima de 0,80 e correlações (de Spearman-Brown) entre 0,35 e 0,5 e significativas ( $p < 0,01$ ) entre medidas em “profundidade” e “largura”, sugerindo existência de validade de critério. Os escores

de utilização ou impacto de treinamento obtidos nas primeiras foram sempre mais elevados do que os obtidos em “largura”, com base em resultados de testes “t” ( $p < 0,01$ ). Seria mesmo de se esperar que os efeitos de treinamento seriam maiores em desempenhos mais diretamente relacionados aos conteúdos que teriam sido ensinados ou desenvolvidos. A transferência de conhecimentos, habilidades e atitudes, ao longo de todo o conjunto de desempenhos do treinando no trabalho, toma muito mais tempo e, em muitos casos, pode nunca ocorrer.

Estratégias de mensuração em “largura” têm também suas limitações. Não cabe ao subsistema de avaliação de treinamentos o papel de definir desempenhos esperados em toda a organização e, se isso for tentado, poderá ser considerado, em muitas culturas organizacionais, como uma intromissão indevida. Além disso, a atual política de flexibilização ocupacional tem colocado à prova as práticas de descrição de cargos e funções. Uma esperança, para os que desejarem desenvolver medidas nesse nível de complexidade, é o movimento de gestão de competências, que tem retomado essas práticas de descrição.

A terceira questão a ser respondida diz respeito a como a medida será realizada: observar ou perguntar? Já foram aqui descritas experiências bem sucedidas utilizando essas duas alternativas, no nível de mensuração de reações. Contudo, no nível de mensuração de comportamento no cargo, a primeira alternativa limita a avaliação ao que pode ser diretamente observado, durante o tempo em que houver observador presente. Isso traz sérias limitações, considerando-se que atualmente a maioria dos treinamentos espera que as pessoas desenvolvam uma ampla variedade de desempenhos complexos e difíceis de serem observados em ambientes e tempos predefinidos (e.g., habilidades para solucionar problemas, estratégias cognitivas e posturas éticas e de civismo organizacionais ou extra-organizacionais). Além disso, na maioria de treinamentos, pode-se esperar mudanças de desempenhos em períodos pós-treinamento muito diversificados, já que os indivíduos são colocados em ambientes com desenhos de trabalho com amplas e elevadas alternâncias ocupacionais, ou em que cada equipe define como e quando cada tarefa será realizada. Talvez por todas essas razões, mais os fatos de que a presença de observadores é vista como ingerência indesejável em muitas culturas organizacionais (especialmente quando os treinamentos são gerenciais ou estratégicos) e que procedimentos de observação requerem atividades de coordenação e treinamento que demandam muitos recursos, a mensuração tem sempre acabado por ser feita na forma de perguntas.

As perguntas podem ser respondidas pelos próprios ex-treinandos (auto-avaliação) ou por seus supervisores, colegas ou clientes (heteroavaliação), ou por uma combinação desses tipos de respondentes. Contudo, se os treinamentos

foram feitos de forma massificada, aos gerentes caberá avaliar simultaneamente o seu impacto em todos os indivíduos de seu setor, o que significará a suspensão ou redução de suas atividades, para poderem responder a tantos questionários. Em culturas organizacionais mais tradicionais, pode não ser bem vista a consulta a colegas ou clientes. A combinação de estratégias seria o ideal, porque permitiria a verificação de diferenças e coincidências de opiniões. Mas é ela que também exige mais recursos, pois demanda grande esforço de coordenação na coleta de dados, para que a correspondência entre eles não seja perdida, no momento da análise. Além disso, também a coleta cruzada de informações pode não ser muito bem vista, se for interpretada como desconfiança do avaliador.

Borges-Andrade e Siri (2000) e Bastos et al. (1999) utilizaram auto e heteroavaliações, com questionários que continham listas específicas de desempenhos esperados dos treinamentos. No primeiro destes estudos, solicitou-se que julgamentos de impactos nesses desempenhos fossem avaliados pelos ex-treinandos (auto-avaliação) e seus supervisores e colegas (heteroavaliações). Foram obtidas estruturas unifatoriais para cada uma das oito dimensões de desempenho investigadas, com alfas de Cronbach variando entre 0,80 e 0,97 nos dados do questionário de auto-avaliação. Entretanto, não foi possível calculá-los para os dados obtidos com o questionário dos colegas e o dos supervisores, pois eles tiveram um baixo percentual de retorno. No segundo estudo, respostas de ex-participantes e de seus chefes foram utilizadas em questionários separados. Mas o número reduzido de casos impossibilitou a verificação da validade de construto e da confiabilidade das medidas. Antes disso, Alves e Tamayo (1993) haviam feito um esforço de construção de um instrumento a ser respondido pelo empregado da Telebrás e outro por seu gerente, mas utilizando um conjunto comum de itens que podiam ser usados para qualquer treinamento. No entanto, apesar de serem minimamente apropriados os resultados das análises fatoriais de cada instrumento, as estruturas fatoriais dos questionários de auto e heteroavaliação diferiram em número e natureza e os índices de confiabilidade (alfas de Cronbach) do primeiro deles foram bastante inferiores aos do segundo. Com isso, as possibilidades de comparação entre escores ficaram limitadas.

Abbad (1999) adaptou seu instrumento para auto-avaliação de impacto de treinamento em qualquer tipo de desempenho, de modo que pudesse também ser enviado para os supervisores de sua amostra de ex-treinandos. Os itens permaneceram os mesmos, mas foram modificados os enunciados e as instruções das escalas de julgamento, de modo que eles pudessem fazer referência a uma terceira pessoa (heteroavaliação), ao invés de uma primeira pessoa (auto-avaliação). Contudo, motivos relacionados a fatos políticos



e aspectos culturais da organização que iria ser estudada impediram a aplicação de seu questionário de heteroavaliação. Esse questionário foi então aplicado entre supervisores de ex-treinandos, juntamente com o de auto-avaliação, em quatro outras organizações, por Borges-Andrade et al. (1999b), Pantoja et al. (1999), Borges-Andrade et al. (1999a) e Pilati et al. (1999). Eles encontraram, nos dados obtidos por ambos instrumentos, características psicométricas (resultantes de análises fatoriais e cálculos de confiabilidade) muito similares às que já tinham sido encontradas por sua autora na versão de auto-avaliação, o que confirmou a validação de construto e a fidedignidade da medida anteriormente desenvolvida. Portanto, isto ampliou sua generalidade. As comparações com testes “t” feitas entre os escores obtidos em auto e heteroavaliação indicam padrões muito distintos ( $p < 0,01$ ), que parecem ser determinados pelo tipo de cultura organizacional e pela natureza da ocupação e dos objetivos dos treinamentos avaliados. Às vezes, os supervisores se mostram mais rigorosos, registrando níveis mais baixos de avaliação nos níveis de comportamento no cargo, outras vezes são os próprios ex-treinandos que se mostram mais exigentes consigo mesmos. Na única vez em que colegas foram envolvidos, seus julgamentos revelaram menores níveis de avaliações de impacto que aqueles dos ex-treinandos e de seus chefes.

É preciso ressaltar que muitos dos estudos aqui relatados, no nível de mensuração de comportamento no cargo, introduziram algumas perguntas abertas, solicitando ao respondente que descrevesse efeitos dos treinamentos no trabalho ou, mais especificamente, que descrevesse impactos positivos e negativos. Nos casos em que o estudo tenha sido feito com listas de desempenhos definidos a partir dos programas de treinamento, geralmente essas questões foram colocadas após cada categoria de efeito esperado. As análises dos conteúdos das respostas quase sempre confirmaram os resultados quantitativos. A interpretação disso e o problema metodológico nela embutido podem ser similares ao que foi apresentado anteriormente, quando se descreveu e discutiu as medidas no nível de reação. Finalmente, e também da mesma forma que neste nível, o balanço que se faz aqui é otimista. Os estudos no nível de comportamento no cargo demonstram que se avançou no sentido de enfrentar os desafios da construção de medidas. Igualmente, as soluções encontradas produziram, elas próprias, outros desafios. Há dilemas que persistirão, pois as soluções tendem a atender alguns condicionantes do processo de tomada de decisão descrito, mas deixam outros a descoberto.

### *Medidas organizacionais*

Ao contrário do nível anterior, no processo de desenvolvimento de medidas de avaliação de treinamento no nível organizacional não existe sistematização metodológica si-

milar para a tomada de decisões. Contudo, pode-se tentar fazer alguns paralelos. O que se deseja efetivamente medir são mudanças que o treinamento possa ter provocado na organização. O foco deve ser no comportamento da organização ou de suas unidades e, contrariando o nível anterior, nunca em seus resultados, pois isso já significaria estar no próximo nível: o de valor final. A medida pode ser feita através de observação (isto inclui documentos) e perguntas. Os dados podem ser fornecidos pela mesma variedade de atores do nível anterior.

As decisões dependerão dos condicionantes já descritos, mas sabe-se que o processo de avaliação seria muito facilitado, se existisse um sistema de controle e avaliação organizacionais em funcionamento e se este incluísse indicadores que fossem minimamente sensíveis para a detecção de efeitos de treinamento. Infelizmente isso raramente acontece, pois esses sistemas (quando existem) não são desenhados com o foco em “T & D”, que para seus construtores é “micro” demais, para ser levado em consideração no nível “macro” em que aqueles sistemas são desenvolvidos. Fazendo uma analogia com o nível anterior, seria como dispor de boas descrições de cargos ou funções, ou de competências, que pudessem servir de parâmetros para se verificar o impacto do treinamento. A diferença é que se estaria esperando por indicadores equivalentes, no nível da organização como um todo ou de suas unidades. Se a presença de tais indicadores, no nível anterior, já não é tão freqüente, talvez fosse otimismo em demasia acreditar que eles pudessem existir e estarem apropriadamente definidos, para servirem ao presente nível de mensuração.

Em princípio, as mudanças de comportamento da organização a serem verificadas deveriam ser derivadas dos desempenhos esperados do conjunto dos indivíduos treinados, que por sua vez deveriam ser deduzidos dos objetivos dos treinamentos. Ao desafio que já existia e foi anteriormente descrito, de fazer essa dedução destes objetivos, acrescenta-se aqui o de derivar mudanças de comportamento da organização a partir de desempenhos individuais esperados, que raramente estão definidos, até porque a tendência de flexibilizar ocupações torna esse trabalho eternamente inacabado. As mudanças nunca deveriam ser modificações no nível do comportamento dos ex-treinandos e poderiam estar localizadas na cultura, em processos ou em estruturas organizacionais. O desafio, neste caso, é de dispor de indicadores apropriados, como já foi mencionado no parágrafo anterior. Há três experiências a relatar.

Borges-Andrade e Siri (2000) e Bastos et al. (1999) usaram os mesmos questionários e respondentes para avaliar treinamentos nos níveis de comportamento no cargo e de organização, mudando somente a descrição do que deveria ser o foco do julgamento. No primeiro estudo, foram listados

principalmente os processos e estruturas que deveriam ser modificados nas instituições, como resultado dos treinamentos em planejamento, controle e avaliação da pesquisa agrícola. Mudanças esperadas na cultura e desempenho organizacionais também foram incluídas. Na segunda pesquisa, os itens cobriram valores e metas organizacionais que, num levantamento exploratório anterior, tinham sido identificados como mudanças organizacionais esperadas do programa concebido para desenvolver habilidades para lidar com o estresse em ambiente hospitalar. Borges-Andrade e Siri (2000) verificaram as diferenças estatísticas, usando ANOVA para medidas repetidas ( $p < 0,01$ ) e testes “post hoc” ( $p < 0,001$ ) entre os julgamentos obtidos no nível de comportamento no cargo e de mudança organizacional. Os escores dos julgamentos referentes a esse primeiro foram sistematicamente mais elevados do que os do segundo, confirmando que os efeitos do treinamento tornam-se menos detectáveis ou confundidos com outros, à medida que se muda de níveis, como prevê a Figura 1.

Borges-Andrade et al. (1999c) tinham como objetivo estudar os efeitos de um treinamento de gerentes de projetos. Para desenvolver sua avaliação no nível organizacional, recorreram a dados que já tinham sido colhidos pela terceira autora, sobre a efetividade de todas as equipes de projetos existentes na Embrapa. Com testes “t”, compararam esses indicadores de efetividade, dividindo as equipes em dois grupos: as que eram coordenadas por ex-treinandos e as que eram coordenadas por indivíduos que não eram ex-treinandos. São dois os problemas metodológicos mais sérios, nesse caso. Os grupos comparados não foram selecionados aleatoriamente e os indicadores de efetividade das equipes não cobriam todos os resultados esperados dos treinamentos dos gerentes. Mas o estudo é um exemplo de como é possível fazer avaliações de treinamento, com a utilização de dados já existentes na organização e com custos muito mais baixos do que os das duas experiências anteriormente mencionadas.

Há muito que avançar, no que tange ao desenvolvimento de medidas no nível organizacional. O caminho a seguir parece ser análogo ao que foi trilhado no nível de comportamento no cargo. Existem desafios a vencer, mas as soluções parecem mais bem dimensionadas, ou existe maior capacidade desenvolvida para superá-los, do que aquilo que foi descrito no nível de mensuração de aprendizagem.

### *Medidas de valor final*

Nos quatro níveis anteriores, a descrição das questões para tomada de decisão e das experiências de mensuração teve como base os modelos e quadros referenciais vindos da Psicologia, Educação e Administração. Os indicadores que têm sido usados para realizar a avaliação do treinamento no nível de valor final, bem como os procedimentos que levam

à sua obtenção, pertencem a uma outra disciplina: a Economia. Eles baseiam-se numa tradição de cálculo de retorno de investimentos que teve seu apogeu há mais de trinta anos, principalmente na Economia da Educação. Como exemplos, ver Becker (1960), Schultz (1961), Mincer (1962), Hansen (1963) e Blaug (1965), embora o trabalho pioneiro seja muito mais antigo (Walsh, 1935). O uso desses métodos foi, em seguida, freqüentemente feito no Brasil, com a exigência dos bancos internacionais de desenvolvimento de que os empréstimos que concediam fossem objeto de avaliações dessa natureza. Textos da época, que podem ser consultados, são os de Patrick e Kehrberg (1973) e Ribeiro (1979).

O primeiro desafio para realizar estudos de avaliação de treinamento no quinto nível é o de contar com profissionais capacitados vindos dessa subárea. O segundo desafio já foi objeto de apreciação aqui, só tornando-se mais complexo, neste último nível: para identificar seus parâmetros de avaliação, é preciso perguntar sobre as mudanças organizacionais e de comportamento individual esperadas do treinamento e, além disso, conhecer os seus objetivos instrucionais. Se não houve planejamentos nos níveis anteriores, como sugere a Figura 1, ou se esse planejamento não foi documentado na organização ou em seu sistema de “T & D”, será preciso recuperar essas informações em todos os níveis anteriores. Definidos os parâmetros específicos da organização e do treinamento, eles certamente envolverão indicadores de custos diretos e indiretos (isto deve incluir até os salários diretos e indiretos dos treinandos, durante o período em que estiveram participando dos eventos) e de benefícios diretos, como a parcela do aumento de produtividade e lucro atribuível ao treinamento, e às vezes indiretos (esses são extremamente difíceis de serem estimados e por isso são muitas vezes ignorados). O problema, em seguida, seria coletar os dados e depois fazer os cálculos, para os quais já foi dito que existe metodologia sistematizada (estimativa de retorno de investimentos) na Economia da Educação.

Raramente se consegue implementar avaliações de treinamento no nível de valor final, apesar dessa metodologia não ser recente. No exterior, há que registrar o esforço realizado por Cascio (1989), embora tenha sido aparentemente uma ação isolada (fora do contexto da estimativa de retorno de investimentos em educação formal). Ávila, Borges-Andrade, Irias e Quirino (1983), adaptando uma metodologia proveniente da mencionada subárea, realizaram um estudo que pode ser considerado como sendo deste último nível de avaliação de treinamento, mas foi igualmente uma ação isolada (embora tenha incluído treinamentos de curta duração e educação formal). Seu objetivo foi o de verificar o retorno dos investimentos em programas de treinamento em capacitação contínua e pós-graduação da Embrapa. Foram utilizados procedimentos de cálculo de rentabilidade social

de investimentos, concentrando-se unicamente em efeitos sociais, verificados em termos de novas tecnologias agrícolas geradas pelos ex-treinandos e adotadas por produtores rurais. Considerou-se ainda, para esse cálculo, todos os custos dos referidos programas.

A motivação para a montagem de uma equipe especial e interdisciplinar (com a presença da Economia, Agronomia, Veterinária, Biologia, Administração, Educação, Sociologia e Psicologia), para realizar esse trabalho, envolveu uma demanda externa de bancos internacionais de desenvolvimento e um desejo especialmente grande da alta direção daquela Empresa. É preciso levar em conta esses fatores condicionantes, antes de uma tomada de decisão semelhante. Outros fatores que devem ser ponderados, pois eles também condicionam a aplicação de uma metodologia dessa natureza, são que não vale a pena fazer esse esforço para investimentos pequenos e para poucos treinamentos. O ideal é envolver todos os programas ou políticas organizacionais em “T & D”. Não é igualmente recomendável que a coleta dos dados seja feita para curtos períodos de tempo de treinamento. Na Embrapa, por exemplo, considerou-se um período médio de sete anos, entre o final dos treinamentos e o início dos lucros advindos das tecnologias adotadas pelos produtores rurais.

Apesar de existir muito pouca experiência realizada no nível de mensuração de valor final, há métodos desenvolvidos e disponíveis para isso. Portanto, a questão aqui não parece ser a de existir uma necessidade de avanço da pesquisa, mas de tomadas de decisão que são condicionadas pelos fatores já citados e que encontram os desafios anteriormente mencionados.

## Conclusão

Os dilemas metodológicos relativos à construção de medidas em avaliação de treinamento foram eleitos como foco do presente texto, partindo-se da premissa de que é legítimo o esforço no sentido de construí-las. A crescente demanda por avaliação, nas organizações que têm sistemas de “T & D”, foi apontada e argumentou-se que as diferenças conceituais entre “T” e “D” estão ficando cada vez mais difíceis de serem encontradas na prática. A necessidade da convivência com diversas estratégias metodológicas de avaliação de treinamento foi reconhecida, mas é preciso conhecer melhor os condicionantes que levam a diferentes escolhas metodológicas. Foram mencionados e discutidos os instrumentos e procedimentos que têm sido usados para fazer avaliações, utilizando-se o modelo de níveis de mensuração de Hamblin (1978), e argumentou-se sobre o avanço alcançado em cada um desses níveis.

No nível de **reação**, houve avanços no sentido de enfrentar os desafios da construção de medidas, existindo sis-

tematização para decidir o que medir, bem como instrumentos já elaborados e testados. As soluções desenvolvidas produziram, elas próprias, outros desafios. Há dilemas que persistirão, pois as alternativas disponíveis tendem a responder algumas questões, mas deixam outras sem meios de serem superadas. No nível de mensuração de **aprendizagem**, ao contrário, não se avançou muito. Os desafios existentes não foram apropriadamente enfrentados, faltando priorizar esforços e realizar pesquisas de forma mais sistemática. No nível de **comportamento no cargo**, há otimismo equivalente ao do primeiro nível, sendo que as pesquisas avançaram no mesmo sentido daquelas do nível de reação. Seus resultados igualmente produziram outros desafios.

O caminho a ser seguido, no nível **organizacional**, tenderá a ser análogo ao anterior, mas ainda há muito que fazer, pois poucas experiências foram realizadas. Essas experiências sugerem que existirão soluções bem diversificadas, embora dependendo em demasia de macro sistemas organizacionais. Há ainda mais escassez de experiências no nível de mensuração de **valor final**, embora neste caso já pareça estar disponível uma opção metodológica. Seus limitantes também se referem à dependência mencionada e à sujeição a vontades externas ao sistema de “T & D” e à própria organização, bem como de existir um programa de treinamento suficientemente grande para justificar o grande esforço que deve ser feito.

Comparando-se os cinco níveis de avaliação de treinamento, pode-se concluir que houve mais avanços metodológicos nos níveis de reação e de comportamento no cargo e uma quase estagnação no nível de aprendizagem. No nível de organização, já se sabe o que deverá ser feito para desenvolver métodos, embora eles ainda estejam raramente disponíveis. Em contraste, no de valor final já se sabe o que fazer e já existem métodos desenvolvidos, mas sua implementação tem sido muito rara. Finalmente, há que destacar que o modelo de Hamblin (1978), cujos componentes foram propostos como níveis hierárquicos de planejamento e de efeitos do treinamento, serve igualmente para sugerir o aumento em complexidade e quantidade dos condicionantes das decisões para construir as medidas desses níveis. Espera-se que o presente texto tenha efetivamente destacado pesquisas divulgadas no Brasil e apontado os seus problemas, de forma a contribuir com futuros estudos e experiências de avaliação de treinamento.

## Agradecimentos

Apoios: CNPq e PRONEX/MCT.

## Referências

Abbad, G., Gama, A. L. G., & Borges-Andrade (2000). Treinamento: análise do relacionamento da avaliação nos níveis de reação, aprendiza-

- gem e impacto no trabalho. *Revista de Administração Contemporânea*, 4(3), 25-45.
- Abbad, G. (1999). *Um modelo integrado de avaliação do impacto do treinamento no trabalho - IMPACT*. Tese de doutorado não-publicada, Universidade de Brasília, Brasília.
- Abbad, G., Sallorenzo, L. H., & Gama, A. L. G. (1998). Treinamento de pessoal: observação do instrutor em sala de aula [Resumo]. In Sociedade Brasileira de Psicologia (Org.), *XXVIII Reunião Anual de Psicologia. Resumos* (p. 204). Ribeirão Preto: Autor.
- Alves, A. R., Pasquali, L., & Pereira, M. A. M. (1999). Escala de Satisfação com o Treinamento – ESAST / TELEBRAS / UnB. *Revista de Administração de Empresas*, 39(1), 25-30.
- Alves, A. R., & Tamayo, A. (1993). Sistema de avaliação do treinamento da Telebrás - SAT. *Revista de Administração*, 28(4), 73-80.
- Ávila, A. F. D., Borges-Andrade, J. E., Irias, L. J. M., & Quirino, T. R. (1983). *Formação do capital humano e retorno dos investimentos em treinamento na Embrapa*. Brasília: Embrapa-DDM/DRH.
- Bastos, A. V. B. (1991). O suporte oferecido pela pesquisa na área de treinamento. *Revista de Administração*, 26(4), 87-102.
- Bastos, A. V. B., Fernandes, S. R. P., & Viana, A. V. (1999). Desenvolvimento de competências e aprendizagem organizacional: avaliação do programa “Cuidar-se para Cuidar” [Texto completo]. In Associação Nacional dos Programas de Pós-graduação em Administração (Org.), *23<sup>o</sup> Encontro Nacional da ANPAD*. (Texto em CD-Rom, p. 1-15). Foz do Iguaçu: Autor.
- Becker, G. S. (1960). Under investment in college education? *The American Economic Review*, 50(2), 346-354.
- Blaug, M. (1965). The rate of return on investment in education in Great Britain. *The Manchester School of Economic and Social Studies*, 33, 205-251.
- Borges-Andrade, J. E. (1982). Avaliação somativa de sistemas instrucionais: integração de três propostas. *Tecnologia Educacional*, 46, 29-39.
- Borges-Andrade, J. E., & Abbad G. (1996). Treinamento no Brasil: reflexões sobre suas pesquisas. *Revista de Administração*, 31(2), 112-125.
- Borges-Andrade, J. E., Azevedo, L. P. S., Pereira, M. H. G. G., Rocha, K. C. P., & Puente K. E. (1999a). Impacto de Treinamentos no Trabalho: o caso do Banco do Brasil [Resumo]. In Sociedade Brasileira de Psicologia (Org.), *XXIX Reunião Anual de Psicologia. Resumos* (pp. 53c-53d). Ribeirão Preto: Autor.
- Borges-Andrade, J. E., Gama, A. L. G., & Oliveira-Simões, J. T. (1999b). Impacto do treinamento no trabalho: um estudo de caso na Elettronorte [Resumo]. In Sociedade Brasileira de Psicologia (Org.), *XXIX Reunião Anual de Psicologia. Resumos* (pp. 53b-53c). Ribeirão Preto: Autor.
- Borges-Andrade, J. E., Morandini, D. C., & Machado, M. S. (1999c). Impacto de treinamento gerencial e efetividade de equipes em ambientes de inovação tecnológica [Resumo]. In Sociedade Brasileira de Psicologia (Org.), *XXIX Reunião Anual de Psicologia. Resumos* (p. 53d). Ribeirão Preto: Autor.
- Borges-Andrade, J. E., & Siri, C. (2000). Impacts of training: the PM&E Project's training. In D. Horton, R. Mackay, A. Andersen & L. Dupleich (Orgs.), *Evaluating capacity development in planning, monitoring, and evaluation: a case from agricultural research* (Study N<sup>o</sup>. 3, CDRom, 1-56). The Hague: International Service for National Agricultural Research.
- Cascio, W. F. (1989). Using Utility Analysis to Assess Training Outcomes. In I. L. Goldstein (Org.), *Training and development in organizations* (pp. 63-88). San Francisco: Jossey Bass.
- Hamblin, A. C. (1978). *Avaliação e controle do treinamento*. São Paulo: McGraw-Hill do Brasil.
- Hansen, W. L. (1963). Total and private rates of returns to investment in schooling. *The Journal of Political Economy*, 71(1), 128-140.
- Leitão, J. S. S. (1994). *Relações entre clima organizacional e transferência de treinamento*. Dissertação de mestrado não-publicada, Universidade de Brasília, Brasília.
- Lima, S. M. V., & Borges-Andrade, J. E. (1985). Meta-análise de avaliação de treinamento. *Revista de Administração*, 20(3), 39-52.
- Lima, S. M. V., Borges-Andrade, J. E., & Vieira, S. B. A. (1989). Cursos de curta duração e desempenho em instituições de pesquisa agrícola. *Revista de Administração*, 24(2), 36-46.
- Martins, M. C. F., Pinto Jr., H., & Borges-Andrade, J. E. (1999). Impacto do treinamento numa empresa de transporte de passageiros [Resumo]. In Sociedade Brasileira de Psicologia (Org.), *XXIX Reunião Anual de Psicologia. Resumos* (p. 53a). Ribeirão Preto: Autor.
- Mincer, J. (1962). On-the-job training: costs, returns and some implications. *The Journal of Political Economy*, 70(5), 50-80 (suplemento).
- Nadler, L. (1984). *The handbook of human resources development*. New York: Wiley.
- Pantoja, M. J., Lima, S. M. V., & Borges-Andrade, J. E. (1999). Avaliação de impacto de treinamento na área de reabilitação: preditores individuais e situacionais [Texto completo]. In Associação Nacional dos Programas de Pós-graduação em Administração (Org.), *23<sup>o</sup> Encontro Nacional da ANPAD*. (Texto em CD-Rom, pp. 1-14). Foz do Iguaçu: Autor.
- Paula, S. M. A. (1992). *Variáveis predictoras de impacto de treinamento no trabalho: análise da percepção dos treinandos de duas organizações*. Dissertação de mestrado não-publicada, Universidade de Brasília, Brasília.
- Patrick, G. F., & Kehrberg, E. W. (1973). Costs and returns of education in five agricultural areas of Eastern Brazil. *American Journal of Agricultural Economics*, 55(2), 145-153
- Pilati, R., Borges-Andrade, J. E., & Azevedo, L. P. S. (1999). Impacto do treinamento em amplitude e profundidade: relações com suporte à transferência, gestão do desempenho e liberdade decisória [Resumo]. In Sociedade Brasileira de Psicologia (Org.), *XXIX Reunião Anual de Psicologia. Resumos* (pp. 53a-53b). Ribeirão Preto: Autor.
- Ribeiro, J. L. (1979). A contribuição da educação na produção agrícola. *Revista de Economia Rural*, 17(4), 85-118.
- Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, 51(1), 1-17.
- Walsh, J. R. (1935). Capital concept applied to man. *The Quarterly Journal of Economics*, 49(1), 255-285.

---

*Notas*

---

- <sup>1</sup> Uma versão anterior do presente texto foi divulgada em CDROM dos Anais do 24<sup>o</sup> Encontro Anual da Associação Nacional dos Programas de Pós-graduação em Administração, Florianópolis, SC, setembro de 2000. Contudo, duas revisões posteriores produziram alterações substanciais no mesmo.
- <sup>2</sup> Desde 1980, eram pessoas ligadas à UnB, Embrapa e Telebrás, responsáveis por "... aproximadamente 75% da produção de conhecimentos e tecnologias na amostra de artigos ..." nacionais analisada por Borges-Andrade e Abbad (1996, p. 116). Incluíam-se professores e estudantes de graduação e pós-graduação em Psicologia da primeira instituição e pesquisadores e técnicos das outras duas. Esse grupo teve seu projeto aprovado pelo Programa de Núcleos de Excelência (PRONEX) do MCT, em 1998, e financiado de 2000 a 2004. O projeto, denominado "Treinamento e Comportamento no Trabalho", é coordenado a partir do Instituto de Psicologia da UnB e também inclui pessoal de Administração e Psicologia da UFBA, Administração da UFMG e Psicologia da UFU.
- 

*Jairo Eduardo Borges-Andrade*, doutor em Sistemas Instrucionais pela The Florida State University, Tallahassee, Florida, EUA, é professor do Departamento de Psicologia Social e Trabalho/Instituto de Psicologia/Universidade de Brasília. Endereço para correspondência: SQN 106 – I – 605; 70742-090, Brasília, D. F. Telefones: (61) 3072625, r. 221; 2724448; 99878259. E-mail: jeborges@linkexpress.com.br.