
Treinamento: Análise do Relacionamento da Avaliação nos Níveis de Reação, Aprendizagem e Impacto no Trabalho

Gardênia Abbad
Ana Lidia Gomes Gama
Jairo Eduardo Borges-Andrade

RESUMO

O presente trabalho tem como objetivo relatar a validação de escalas de avaliação de **reações** ao treinamento, bem como os resultados da análise do relacionamento entre os três mais tradicionais níveis: **reação**, **aprendizagem** e **impacto do treinamento no trabalho**. A literatura especializada tem mostrado que os relacionamentos entre esses níveis nem sempre são significativos ou estão na direção tradicionalmente prevista. Apesar de muito importantes, medidas válidas e confiáveis de reação ainda são raras. O questionário de avaliação de reações descrito nesse trabalho contém 39 itens. Foi aplicado em uma amostra de servidores públicos participantes de 229 cursos de curta duração. As aproximadamente 3.700 respostas válidas aos questionários foram submetidas à análise dos principais componentes e à análise fatorial (PAF, rotação oblíqua - *direct oblimin*), e à análise de consistência interna (alfa de Cronbach). Foram obtidas duas escalas altamente confiáveis: uma concernente às reações ao treinamento e outra referente ao desempenho do instrutor. Quanto ao relacionamento entre os níveis de avaliação, encontrou-se que **reações** estão fortemente correlacionadas com **impacto**, porém fracamente com **aprendizagem**. Os resultados são discutidos em termos de suas implicações práticas e teóricas para avaliação de treinamento.

Palavras-chaves: treinamento; aprendizagem; reações ao treinamento; impacto do treinamento; treinamento e desempenho no trabalho.

ABSTRACT

The present paper aims to report the validation of **training reaction** scales, as well as to report the results from an analysis of the relationship among the three most common levels: **reaction**, **learning** and **impact of training on work**. The scientific literature has shown that the relationship between these levels has not always been significant, or toward the predicted direction. Despite their relevance, valid and reliable reaction measures are very rare. The used reaction questionnaire contains 39 items. It was applied in a sample of 229 training events. Approximately 3,700 valid questionnaires were subjected to factorial analysis (PAF – direct oblimin) and internal consistency analysis (Cronbach’s alpha). Two scales with high level of reliability were obtained: one of them concerning training reactions and the another concerning the instructor performance. Finally, very strong relationships were found between the **reaction** and **impact** levels, but poor relationships between the **reaction** and the **learning** level. The results are discussed in terms of their practical and theoretical implications for training evaluations.

Key words: training; learning; reactions to training; impact of training; training and performance at work.

INTRODUÇÃO

O presente artigo tem como objetivos principais descrever os procedimentos de validação de escalas de avaliação de **reações** ao treinamento, bem como uma análise do relacionamento entre os três mais tradicionais níveis de avaliação: **reação**, **aprendizagem** e **impacto de treinamento no trabalho**. Entende-se por **reações** o nível de satisfação dos participantes com a programação, o apoio ao desenvolvimento do curso, a aplicabilidade, a utilidade e os resultados do treinamento. As escalas descritas neste trabalho foram desenvolvidas principalmente com base no Modelo de Avaliação Somativa de Sistemas Instrucionais – MAIS (Borges-Andrade, 1982) e no trabalho de Lima e Borges-Andrade (1985). Estas também fazem parte do conjunto de instrumentos de coleta de dados utilizados por Abbad (1999) na pesquisa que avaliou 229 cursos de curta duração nos níveis de reação, aprendizagem e impacto do treinamento no trabalho. **Aprendizagem**, neste estudo, corresponde ao quinto componente do modelo IMPACT de Abbad (1999) e refere-se ao grau de assimilação dos conteúdos ensinados no treinamento, medido em termos de escores obtidos pelo participante em testes ou provas aplicadas pelo instrutor no final do curso. A variável **impacto do treinamento no trabalho** é aqui definida como a auto-avaliação feita pelo próprio participante acerca dos efeitos mediatos produzidos pelo treinamento em seus níveis de desempenho, motivação, auto-confiança e abertura a mudanças nos processos de trabalho. Reconhece-se que essa avaliação também pode ser feita por outras pessoas, como o supervisor, colegas, ou clientes daquele participante.

Apesar de muito importantes, medidas válidas e confiáveis de satisfação com o treinamento são ainda raras. No Brasil, conforme a revisão de Borges-Andrade e Abbad (1996), os artigos sobre avaliação de treinamento, em sua maioria, fazem referência apenas à avaliação de aprendizagem. Pesquisadores brasileiros, entretanto, têm utilizado em seus trabalhos, algumas medidas de reação, tais como Quirino, Borges-Andrade e Pereira (1980), Lima e Borges-Andrade (1984), Borges-Andrade, Lima e Reis (1989), Lima (1989), Alves e Tamayo (1993) e Alves, Pasquali e Pereira (1999).

A literatura especializada em T&D vem mostrando resultados inconsistentes no que tange ao relacionamento de variáveis de reações ao treinamento com medidas tradicionais como **aprendizagem** e **comportamento no cargo (impacto)**, em parte devido à falta de medidas psicometricamente confiáveis, desenvolvidas com base em modelos teóricos consistentes.

Conforme Alliger e Janak (1989), os relacionamentos entre aprendizagem, reações e impacto no cargo nem sempre são significativos ou estão na direção prevista pelas abordagens mais tradicionais de avaliação. Os resultados de algumas pesquisas têm descrito situações nas quais um participante, apesar de demonstrar satisfação (reação favorável) com o treinamento e de obter bons escores nas avaliações de aprendizagem, não transfere para o trabalho as novas habilidades aprendidas no treinamento.

As abordagens mais utilizadas na atualidade, a de Kirpatrick (1976, 1977) e a de Hamblin (1978), sugerem que as variáveis-critério reações, aprendizagem, desempenho no cargo e resultados (ou mudança organizacional e valor final) mantêm entre si um forte relacionamento positivo. A taxonomia de níveis de avaliação sugerida por Kirpatrick (1976, 1977) tem sido a mais utilizada nas pesquisas estrangeiras sobre avaliação de treinamento, mesmo após transcorridos muitos anos desde a sua concepção. Segundo Alliger e Janak (1989), na proposta original de Kirpatrick (1976) os quatro níveis estão articulados entre si em uma seqüência hierárquica de produção de informações sobre a eficácia de um treinamento. Isto de tal forma que o nível mais baixo de informação corresponde à medida de reação e o mais alto, à medida de resultados. Alliger e Janak (1989), no entanto, têm considerado problemáticos os três pressupostos básicos do modelo de Kirpatrick (1976), nos quais os níveis, arranjados em uma ordem crescente de informação, encontram-se ligados uns aos outros por elos de causalidade e de correlações positivas. O primeiro pressuposto, segundo tais autores, poderia levar o pesquisador a crer que o nível resultados é invariavelmente melhor que os demais, o que seria um equívoco. Isso seria falso, por exemplo, quando não interessasse à organização provocar mudanças em todos os níveis. Quanto ao segundo e terceiro pressupostos, eles não têm recebido confirmação empírica. Sob o ponto de vista lógico, seria um equívoco, também, por exemplo, inferir causalidade entre reação e aprendizagem, quando, muitas vezes, ambas as medidas são tomadas ao mesmo tempo. Além disto, ao contrário do que supunha Kirpatrick (1976), os estudos revisados por Alliger e Janak (1989) mostraram correlações negativas entre reação e aprendizagem.

Na literatura examinada por Alliger e Janak (1989), foram encontrados poucos estudos sobre o relacionamento entre os quatro níveis, o que, para os autores, pode significar que os pesquisadores simplesmente assumem como verdadeiras as afirmações de Kirpatrick (1976) a respeito da existência de correlações positivas entre os critérios, deixando de investigar tais relacionamentos. No Brasil, a situação não parecia muito diferente, quando Borges-Andrade e Abbad (1996) analisaram a literatura sobre T&D.

A meta-análise dos resultados das pesquisas analisadas por Alliger e Janak (1989) mostrou isto: o nível de reação tende a estar fracamente relacionado aos demais critérios, havendo, também, correlações baixas entre aprendizagem, comportamento no cargo e resultados (mudança organizacional e valor final). Os resultados de Abbad (1999) apontaram em outra direção, ao encontrar correlações positivas e estatisticamente significativas entre **reações** e **impacto de treinamento no trabalho**. Os dados sobre o relacionamento entre os níveis de avaliação são contraditórios e muitas vezes inconsistentes. Apesar disso, a maior parte das pesquisas desta área continua sendo concebida com base nas abordagens tradicionais, sem um questionamento sobre a adequação de alguns dos pressupostos desses modelos.

Goldstein (1991) acredita que as medidas de reação são pouco úteis como preditoras de aprendizagem; só têm sentido, se estiverem relacionadas às necessidades de treinamento. Além disso, a aprendizagem deveria ser considerada apenas condição necessária, mas não suficiente, para a transferência ou impacto no trabalho, segundo Goldstein (1991) e Tannenbaum e Yukl (1992).

Os resultados de pesquisas realizadas recentemente e apresentadas por Tannenbaum e Yukl (1992) também não confirmaram o relacionamento significativo entre as medidas de aprendizagem e de transferência (impacto) e entre medidas de reação e os demais critérios (aprendizagem, impacto e resultados). Alguns resultados mostraram, todavia, que o nível de reação exerceu um papel moderador no relacionamento de motivação com aprendizagem. A eficácia de programas instrucionais, tendo em vista o exposto, não deveria ser aferida avaliando-se apenas um nível de variável, como geralmente vem sendo feito nessa área.

Essas descobertas recentes concernentes às fracas relações entre reação, aprendizagem e impacto não são conclusivas, havendo, ainda, poucas tentativas de identificação das variáveis preditoras de cada um desses níveis de avaliação. Um dos motivos para a inconsistência nos resultados das pesquisas é, provavelmente, a falta de instrumentos válidos, confiáveis e comparáveis, construídos a partir de modelos teóricos abrangentes que incluam variáveis preditoras para cada um dos níveis de avaliação. O presente trabalho é uma tentativa desse tipo. As escalas de avaliação de reações descritas a seguir foram construídas com base no modelo MAIS de Borges-Andrade (1982), que vem possibilitando a criação de medidas úteis e abrangentes do nível de satisfação do treinando com sistemas instrucionais.

Nos últimos anos, tem crescido o interesse nacional e internacional pela pesquisa sobre avaliação de treinamento. A área está ativa e vem caracterizando-se pela busca incessante de melhores modelos conceituais, teóricos e metodológicos para estudar as variáveis de interesse.

No Brasil, os dados da análise de literatura de Borges-Andrade e Abbad (1996) são encorajadores. Esses autores encontraram artigos que enfocam modelos, instrumentos, procedimentos e resultados de avaliação de treinamento. Essa tem sido a linha nacional de pesquisa mais bem sucedida na produção de conhecimentos e tecnologias sobre treinamento. O modelo de avaliação somativa de produtos instrucionais (MAIS), proposto por Borges-Andrade (1982), com base no modelo CIPP (contexto, insumo, processo e produto) de Stufflebeam (1984) e na abordagem de sistemas, vem inspirando pesquisadores brasileiros a produzir mais conhecimentos e tecnologias de avaliação de treinamento.

O modelo MAIS foi utilizado, por exemplo, como referencial para os trabalhos de Lima e Borges-Andrade (1984), Alves, Pasquali e Pereira (1999), bem como para o de desenvolvimento de um sistema de avaliação de treinamento, o SAT da Telebrás (Alves e Tamayo, 1993).

As escalas descritas neste trabalho contêm itens relacionados a insumos, procedimentos, processos instrucionais, resultados e ambiente tal como definidos por Borges-Andrade (1982). Insumo refere-se, neste trabalho, às características demográficas e funcionais da clientela dos treinamentos, que podem alterar aspectos desses treinamentos ou de seus resultados. Os procedimentos referem-se à qualidade do desempenho didático do instrutor ao utilizar estratégias e recursos instrucionais em sala de aula, bem como à aspectos de planejamento instrucional. Os resultados dizem respeito aos efeitos imediatos do treinamento e geralmente são operacionalizados em termos de indicadores de aprendizagem. Apoio, neste estudo, corresponde ao suporte material oferecido pela organização para a execução do treinamento, bem como às expectativas do treinando de encontrar no ambiente de trabalho clima propício à transferência de aprendizagem.

Nas seções seguintes são apresentadas as etapas de construção e validação das escalas de **reação**, bem como a análise das relações dessas variáveis com **aprendizagem, impacto de treinamento no trabalho e retenção**.

MÉTODO - CONTEXTO DA PESQUISA

Este trabalho foi desenvolvido no período de março de 1997 a junho de 1998 no Instituto Serzedello Corrêa (ISC), órgão do Tribunal de Contas da União (TCU), com sede em Brasília, DF. As escalas descritas nesse trabalho fazem parte do conjunto de instrumentos de coleta de dados utilizado por Abbad (1999), em uma pesquisa de testagem empírica de um modelo integrado de avaliação de

impacto de treinamento no trabalho, que contemplou os níveis de **reação**, **aprendizagem** e **impacto no trabalho**.

Construção e Validação Semântica das Escalas

Os conceitos que fundamentaram a construção das escalas de avaliação de reação inserem-se principalmente no modelo de avaliação somativa de sistemas instrucionais (MAIS), proposto por Borges-Andrade (1982) que, por sua vez, deriva dos modelos propostos por Kirpatrick (1976, 1977), Hamblin (1978) e Stufflebeam (1984). O sistema de avaliação de treinamento (SAT) de Alves e Tamayo (1993) e os trabalhos de Lima e Borges-Andrade (1985) e Paula (1992) também serviram de base para a elaboração deste estudo.

Reação corresponde ao quarto componente do IMPACT (Abbad, 1999) e refere-se à satisfação do participante com a qualidade do curso e do desempenho do instrutor. O questionário de avaliação de reações, desenvolvido com base no suporte teórico supracitado, contém 39 itens, associados a uma escala de seis alternativas de respostas, a saber: Ótimo (5), Muito Bom (4), Bom (3), Regular (2), Ruim (1) e Não se Aplica (0). Esse instrumento foi desenvolvido a partir de outro já utilizado por Lima e Borges-Andrade (1985). Os 39 itens encontram-se distribuídos em duas partes. A primeira refere-se à avaliação de aspectos concernentes ao curso propriamente dito, que contém cinco dimensões: (1) Programação, representada por 7 itens; (2) Apoio ao Desenvolvimento do Curso, 2 itens; (3) Aplicabilidade e Utilidade do Treinamento, 3 itens; (4) Resultados do Treinamento, 7 itens e (5) Suporte Organizacional, 3 itens.

Os itens concernentes à programação compreendem a opinião do participante acerca da qualidade do curso, em termos de clareza na definição de objetivos, compatibilidade destes com as necessidades de treinamento, carga horária (diária, total e destinada a atividades teóricas e práticas), ordenação e adequação do conteúdo programático aos objetivos do curso. A definição de programação corresponde, em parte, ao componente **procedimento** do modelo MAIS proposto por Borges-Andrade (1982) e de **planejamento** em Lima e Borges-Andrade (1985).

Os itens de apoio ao desenvolvimento do curso referem-se à opinião do participante do treinamento acerca da qualidade das instalações (sala de aula), bem como à qualidade, organização e quantidade do material didático distribuído durante o curso. A definição dessa subdimensão é parte do conceito de **apoio** do modelo MAIS (Borges-Andrade, 1982).

O conjunto de variáveis denominado aplicabilidade e utilidade do treinamento representa a opinião do participante sobre o valor instrumental do treinamento

na resolução de problemas, a possibilidade de aplicação no trabalho dos conhecimentos adquiridos no treinamento, assim como a conveniência da disseminação dos conteúdos do curso para outros servidores.

Os itens relacionados aos resultados referem-se à opinião do participante acerca da efetividade do curso, em termos do quanto foi capaz de assimilar os conhecimentos, capacitação para o reconhecimento de situações onde é correto aplicar as novas habilidades, chances de melhorar o desempenho e a integração do participante com outros profissionais. Além disso, no instrumento de coleta de dados, constam itens acerca do estímulo dado pelo treinamento ao participante para aplicar no trabalho os conhecimentos recém-aprendidos. Essas variáveis fazem parte do conceito **resultados** no modelo MAIS de Borges-Andrade (1982).

Os itens de suporte organizacional exprimem as expectativas do participante acerca do apoio organizacional à transferência de aprendizagem, no que se refere à disponibilidade de recursos, oportunidades e clima propício ao uso das novas habilidades. Essas variáveis foram escolhidas com base nas definições de **clima para transferência** de Roullier e Goldstein (1993), **ambiente de tarefa** (Abbad, Lima e Veiga, 1996), **suporte à transferência** (Broad, 1982) e mantêm um laço teórico com o componente **apoio** do modelo MAIS de Borges-Andrade (1982).

A segunda parte do instrumento contém 15 itens de avaliação da satisfação dos treinandos com o desempenho do instrutor. Os itens encontram-se agrupados em três dimensões: (1) Desempenho Didático, 11 itens; (2) Domínio do Conteúdo, 2 itens e (3) Entrosamento com os Treinandos, 3 itens.

Desempenho didático relaciona-se às avaliações feitas pelos participantes do treinamento acerca da qualidade dos procedimentos e processos instrucionais adotados pelo instrutor em sala de aula. São eles: transmissão de objetivos, resumo e revisão dos temas, nível de profundidade com que os assuntos foram abordados, ritmo de apresentação, uso de estratégias motivacionais, qualidade da apresentação dos conteúdos e exemplos, uso de estratégias e recursos instrucionais, além da qualidade das avaliações de aprendizagem. Essas variáveis foram escolhidas com base, principalmente, nos trabalhos de Driskell, Willis e Cooper (1992), Reiser e Dick (1996) e VanLehn (1996).

Domínio do conteúdo refere-se à opinião do participante sobre o nível de conhecimentos e o grau de segurança demonstrados pelo instrutor, ao abordar assuntos relacionados ao conteúdo programático do curso. Os itens de entrosamento com os treinandos representam a opinião do participante sobre a disposição demonstrada pelo instrutor para esclarecer dúvidas e estimular a manifestação de idéias sobre a matéria ensinada no curso.

Uma vez formatado, o questionário passou por validação semântica, da qual participaram 15 servidores, de nível superior de instrução, pertencentes à equipe técnica e gerencial do Instituto Serzedello Corrêa. Esses servidores, após terem participado de um curso de 40 horas sobre sistemas instrucionais e avaliação de treinamento, julgaram a clareza dos itens, sua adequação e objetividade. Após algumas modificações nas instruções e no formato do instrumento, o questionário foi submetido à validação estatística, como descrito a seguir.

Amostra de Cursos

Este trabalho abrangeu as avaliações de reação feitas por participantes de 229 cursos oferecidos pelo ISC. Desse total, 29,7% eram, conforme classificação elaborada pelo ISC, de conteúdo técnico/administrativo e os demais eram instrumentais. Entre os primeiros, estavam 25 cursos de Auditoria, 20 de Estudos Jurídicos, 10 de Administração, 6 de Atividades Administrativas, 3 de Estudos Internacionais, 2 de Administração Empresarial e 2 para chefes de unidades administrativas. Entre os cursos de natureza instrumental, estavam 80 de Informática, 66 de Gerenciamento Organizacional e 15 de Redação e Gramática. Conforme a taxonomia de Gagné (1968), 70,7% dos cursos avaliados incluíram habilidades de natureza cognitiva e apenas 29,3% incluíram atitudes.

Os cursos estudados contavam com um número médio de 18,11 (DP= 5,70) participantes, sendo que as turmas menores eram compostas por 6 e as maiores por 35 servidores. Na maior parte dos cursos (94,3%), a carga horária total média era de 17,9 horas (DP= 5,89) e a carga diária de 3,08 horas (DP= 0,36).

Amostra de Participantes

O questionário de avaliação de reações foi aplicado em todos os servidores que concluíram os cursos anteriormente descritos. O arquivo de dados desta pesquisa contém um total de 4.051 casos de participação de servidores em treinamento; contudo grande parte dos servidores participaram de mais de um curso durante o período de coleta de dados, de modo que um mesmo participante pode ter avaliado mais de um treinamento. Do total de participantes dos cursos, 71,6% tinham nível superior completo e 72,3% enquadravam-se na faixa etária de 31 a 50 anos. Havia ligeira predominância do gênero masculino (55,5%) e a grande maioria pertencia ao próprio TCU (72%). Entre esses participantes, estavam servidores públicos provenientes de outras 75 instituições públicas estaduais e federais. Ainda sobre as características da clientela, notou-se que mais da metade dos participantes (53,2%) possuía algum cargo de chefia.

Procedimentos

A aplicação do instrumento de avaliação de **reações**, realizada no final de cada curso, ocorria, sempre que possível, antes da aplicação do teste de aprendizagem pelo instrutor (pós-teste de conhecimentos), na tentativa de evitar a perda de respondentes. O questionário foi aplicado coletivamente, na própria sala de aula, por alunos de graduação do curso de psicologia da Universidade de Brasília, que ficavam encarregados de orientar os participantes quanto ao preenchimento correto do instrumento e zelar pela guarda e confidencialidade dos dados.

Os participantes da amostra, ao responderem ao questionário de avaliação de **reações**, já tinham participado anteriormente de duas fases da pesquisa, respondendo ao instrumento de Avaliação de Suporte Organizacional ao Treinamento e resolvendo pré-testes de conhecimentos relacionados aos conteúdos dos cursos, conforme previa o delineamento de avaliação do modelo IMPACT (Abbad, 1999).

Análise de Dados

Os dados obtidos por intermédio dos questionários foram inicialmente submetidos a análises estatísticas descritivas e de cunho exploratório, conforme recomendam Tabachnick e Fidell (1996), objetivando avaliar a correção do arquivo de dados, a distribuição dos dados omissos, as variáveis com distribuição não normal, bem como localizar *outliers*.

Em uma segunda etapa, os dados coletados mediante a aplicação dos questionários foram submetidos à análise dos principais componentes e à análise fatorial (PAF, rotação oblíqua - *direct oblimin*), bem como à análise de consistência interna (alfa de Cronbach). Essas análises permitiram examinar a estrutura empírica dos dados, validar estatisticamente o conteúdo dos instrumentos e julgar a confiabilidade deles.

Todas as análises foram realizadas por intermédio do pacote estatístico SPSS, versão 7.5. A confecção de tabelas e gráficos foi realizada com o apoio dos programas *Excel*, versão 97 e *Word*, versão 7.0. Além desses, foi necessária a utilização do *Access*, versão 97, para indexação de arquivos e a migração de dados cadastrais dos participantes da amostra, fornecidos pelo ISC, referentes à matrícula, lotação, cargo e função para os arquivos de dados.

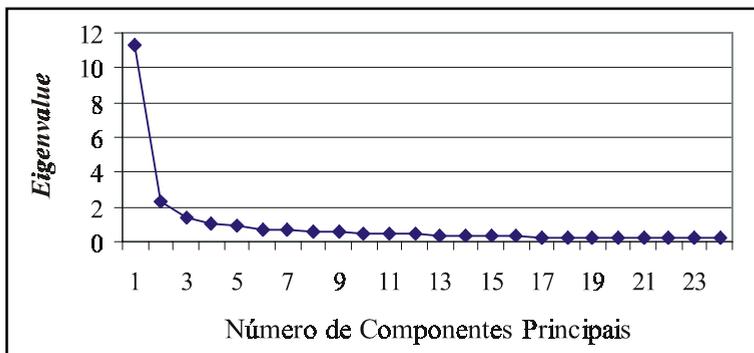
RESULTADOS

Nesta seção, apresentam-se os resultados do processo de validação estatística dos instrumentos. Inicialmente, são descritos os resultados referentes à escala de **reação** aos cursos e, posteriormente, à de **reação** ao desempenho do instrutor.

Escala de Avaliação de Reação aos Cursos

As 3.670 respostas válidas da escala de reações foram submetidas ao processo de validação estatística e de confiabilidade. A análise dos componentes principais mostrou que a matriz de dados era altamente fatorável ($KMO=0,96$) e indicou quatro fatores que explicavam 67,09% da variância total das respostas. A Figura 1 mostra o *scree plot* resultante dessa análise.

Figura 1: Número de Componentes Principais de Reações ao Curso



A inspeção do *scree plot* (Figura 1) e a análise semântica dos componentes dessas escalas mostraram, entretanto, que havia apenas dois fatores conceitualmente precisos. Por esse motivo, os dados foram reanalisados por meio do método dos eixos principais (PAF), forçando-se a solução em dois fatores.

Como resultado dessa segunda análise, foram encontrados dois fatores altamente correlacionados entre si ($r=0,63$), denominados **reações aos resultados, aplicabilidade e expectativas de suporte** e **reações à programação e ao apoio ao desenvolvimento do curso**. As Tabelas 1 e 2 mostram a estrutura empírica dessas subescalas, bem como os valores das comunalidades, cargas fatoriais e índices de confiabilidade.

Tabela 1: Estrutura Empírica do Fator 1 – Reações aos Resultados, Aplicabilidade e Expectativas de Suporte

Descrição dos Itens	H²	Carga
Intenção de aplicar os conhecimentos no trabalho.	0,62	0,89
Probabilidade de melhorar o desempenho do grupo de trabalho.	0,65	0,86
Oportunidades de aplicar no trabalho.	0,53	0,86
Estímulo para aplicar conhecimentos no trabalho.	0,68	0,84
Probabilidade de melhorar o desempenho.	0,67	0,84
Possibilidade de aplicação no trabalho.	0,60	0,82
Clima propício.	0,42	0,78
Reconhecimento correto das situações de aplicação.	0,61	0,71
Possibilitar integração com outros profissionais.	0,50	0,66
Utilidade.	0,64	0,64
Capacidade de transmitir os conhecimentos a outros.	0,43	0,60
Suporte material à transferência de treinamento.	0,36	0,57
Conveniência da disseminação para outros servidores.	0,44	0,54
Assimilação dos conhecimentos.	0,55	0,49

α Fator1 = 0,95

Tabela 2: Estrutura Empírica do Fator 2 – Reação à Programação e ao Apoio

Descrição dos Itens	H²	Carga
1. Ordenação do conteúdo.	0,48	0,84
2. Adequação dos conteúdos aos objetivos.	0,54	0,79
3. Definição de objetivos.	0,45	0,70
4. Carga horária total.	0,29	0,67
5. Qualidade e organização do material didático.	0,30	0,65
6. Quantidade de material didático.	0,29	0,64
7. Carga horária diária.	0,27	0,62
8. Carga horária para atividades práticas.	0,27	0,59
9. Compatibilidade entre objetivos e necessidades.	0,55	0,53
10. Qualidade das instalações.	0,15	0,40

α Fator2 = 0,89

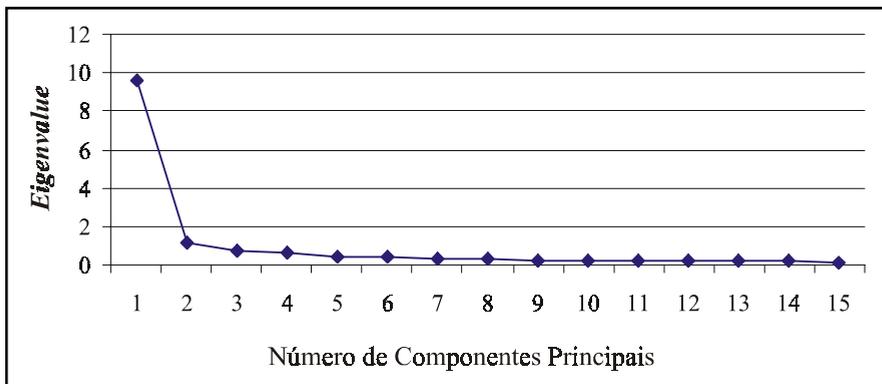
Nota-se, nas Tabelas 1 e 2, que ambas as escalas obtiveram ótimos índices de confiabilidade. Da primeira, constam 15 itens relacionados à satisfação do participante com os efeitos do curso sobre o seu próprio desempenho, bem como às

suas expectativas de aplicação das novas habilidades no trabalho. A segunda subescala contém 10 itens que medem a satisfação do participante com a qualidade do programa, recursos de apoio à instrução e instalações da sala de aula.

Escala de Avaliação do Desempenho do Instrutor

As 3.670 respostas válidas dos itens concernentes à avaliação de desempenho do instrutor foram submetidas à análise dos componentes principais. A matriz de dados mostrou-se altamente fatorável ($KMO=0,96$). Essas análises apontaram uma estrutura com, no máximo, dois fatores que explicavam 71,98% da variância total das respostas. Esses componentes eram altamente correlacionados entre si ($r=0,64$). A inspeção da forma do *scree plot* (Figura 2), bem como a análise dos conteúdos dessas escalas sugeriram uma estrutura unifatorial predominante.

Figura 2: Número de Componentes Principais de Reações ao Desempenho do Instrutor



Por esse motivo, os dados foram submetidos a uma análise fatorial (PAF), forçando-se a solução em um único fator que explicava 61,64% da variância total das respostas. Esse fator, denominado **reação ao desempenho do instrutor**, contém todos os 15 itens do questionário referentes a esse tópico. Nota-se, ainda, que a escala única atingiu altos índices de confiabilidade, tal como descrito na Tabela 3. Esse fator único refere-se à satisfação do participante com a didática, segurança, domínio do conteúdo e competência interpessoal demonstrados pelo instrutor em sala de aula.

Em uma segunda etapa, analisou-se a matriz de dados, forçando-se a solução da análise (PAF) em dois fatores, tal como sugerido pelo formato do *scree plot* (vide Figura 2). Esses dois componentes explicavam 67,95% da variância total

das respostas e estavam fortemente correlacionados entre si ($r=0,74$). A primeira escala (com $\alpha=0,95$) refere-se ao desempenho didático do instrutor e é composta por 10 itens (1, 2, 3, 4, 5, 6, 7, 9, 15, e 14), enquanto a segunda (com $\alpha=0,92$) é composta por 5 itens (8, 10, 11, 12 e 13) (vide Tabela 3).

Tabela 3: Estrutura Empírica do Fator Reação ao Desempenho do Instrutor

Descrição dos Itens	H ²	Carga
1. Qualidade da apresentação dos conteúdos.	0,77	0,87
2. Organização da apresentação.	0,73	0,84
3. Uso de estratégias para motivar o treinando.	0,72	0,84
4. Transmissão de objetivos instrucionais.	0,72	0,84
5. Sumarização dos conteúdos.	0,69	0,82
6. Ritmo de apresentação dos conteúdos.	0,68	0,81
7. Profundidade.	0,67	0,81
8. Segurança.	0,67	0,80
9. Uso de estratégias instrucionais.	0,63	0,77
10. Conhecimento dos temas.	0,62	0,77
11. Esclarecimento de dúvidas.	0,60	0,76
12. Estímulo à participação.	0,58	0,74
13. Respeito às idéias manifestadas pelo treinando.	0,55	0,72
14. Uso de recursos instrucionais.	0,54	0,71
15. Qualidade das avaliações de aprendizagem .	0,43	0,63

Fonte: Abbad, Sallorenzo e Gama (1998).

$\alpha = 0,96$

Ambas as soluções indicam que tanto a escala unidimensional quanto a bidimensional são igualmente úteis, confiáveis e válidas, podendo ser utilizadas juntas ou separadas, dependendo dos objetivos da avaliação. No caso, por exemplo, de utilização dos resultados da avaliação para fornecimento de *feedback* aos instrutores, a estrutura mais apropriada parece ser a de dois fatores, pois possibilitará maior riqueza de informações.

Relações entre os Níveis de Avaliação de Treinamento

Nesta seção, apresentam-se as correlações bivariadas entre os três mais tradicionais níveis de avaliação (**aprendizagem, reações e impacto no trabalho**). Além disso, são feitas avaliações da utilidade e aplicabilidade das escalas de reação no contexto das pesquisas de avaliação de treinamento.

Na Tabela 4 encontram-se os coeficientes de correlação bivariada das variáveis: **aprendizagem**, definida como os escores obtidos por 3.082 participantes nos pós-testes de assimilação dos conteúdos dos cursos; **impacto do treinamento no trabalho**, definida como a média das avaliações feitas pelos próprios participantes, acerca dos efeitos produzidos pelo treinamento em seus níveis de desempenho, motivação, autoconfiança e abertura a mudanças nos processos de trabalho, medido duas semanas após o treinamento, por meio dos itens do fator Impacto A e três meses após o treinamento, por meio dos itens do fator Impacto B. No caso do Impacto A, foram obtidas 1.270 respostas válidas, enquanto do Impacto B, apenas 600. Essas respostas foram fornecidas por uma amostra de servidores do Tribunal de Contas da União, participantes de treinamentos oferecidos pelo ISC, avaliados por Abbad (1999).

Ainda na Tabela 4, o nível de reações encontra-se representado pelas médias das respostas dos participantes aos itens componentes dos fatores **reações à programação e ao apoio** (Reapro), **reações aos resultados, aplicabilidade e expectativas de suporte** (Reares) e **reações ao desempenho do instrutor** (Reains), descritos nas Tabelas 1, 2 e 3.

Além dessas, consta da Tabela 4 a variável **retenção**, medida em termos da avaliação feita pelo participante sobre seu próprio nível de recordação dos conteúdos aprendidos no curso, após transcorridas duas semanas do término do treinamento (Retenção A) e três meses após o treinamento (Retenção B). Nas análises descritas na Tabela 4, foram consideradas aproximadamente 1.270 respostas válidas para Retenção A e 600 para B (ver Abbad, 1999).

Tabela 4: Correlações entre as Variáveis de Resultados Mediatos e Imediatos do Treinamento

Variáveis	Aprendi- zagem	Impacto A	Impacto B	Reapro	Reares	Reains	Reten- ção A	Reten- ção B
Aprendizagem								
Impacto A	0,00							
Impacto B	-0,04	0,76**						
Reapro	0,05**	0,39**	0,31**					
Reares	0,08**	0,59**	0,53**	0,64**				
Reains	0,05**	0,38**	0,35**	0,71**	0,61**			
Retenção A	0,07*	0,48**	0,34**	0,34**	0,37**	0,29**		
Retenção B	0,02	0,43**	0,57**	0,32**	0,40**	0,29**	0,49**	

**significativa no nível de 0,01 (2-tailed) * significativa no nível de 0,05 (2-tailed)
correlação < 0,005 = 0,00

As medidas de **impacto**, **reação** e **retenção**, de modo geral, estão moderadamente relacionadas entre si. A **aprendizagem** é a única variável que não está significativamente correlacionada com todas as demais. Na Tabela 4 nota-se que a variável **reações aos resultados, aplicabilidade e expectativas de suporte** (Reares), entre as variáveis estudadas, é a que está mais fortemente correlacionada com as medidas de **impacto** (A e B). Esses resultados são confirmados por outros apresentados no modelo de regressão múltipla *stepwise* de Impacto A como variável critério, descrito no trabalho de Abbad (1999), onde se observa que quase todas as variáveis preditoras pertencentes ao componente reações (probabilidade de melhorar o desempenho, estímulo para aplicar as novas habilidades, expectativa de clima propício à transferência de treinamento e conveniência da disseminação do curso para outros servidores) são integrantes do referido fator (Reares).

Esses resultados são bastante claros. Há forte relacionamento entre reações e impacto do treinamento no trabalho; porém uma avaliação adequada da eficácia do treinamento a longo prazo não se pode restringir à avaliação das reações, porque há fortes preditoras de impacto (e retenção) que pertencem a outros componentes como, **suporte organizacional** e **características da clientela**, como demonstrou Abbad (1999). Além disso, segundo esta autora, há importantes variáveis explicativas de impacto que se referem ao ambiente pós-treinamento, as quais são, provavelmente, mais bem avaliadas pelo participante em seu retorno ao ambiente de trabalho.

DISCUSSÃO E CONSIDERAÇÕES FINAIS

Esses resultados e aqueles relatados por Abbad (1999) não confirmaram alguns dos principais pressupostos de Kirpatrick (1976, 1977) e Hamblin (1978), de que os níveis de avaliação mantêm entre si forte relacionamento positivo. Isso não foi observado, pelo menos no relacionamento dos três primeiros níveis.

Aprendizagem, por exemplo, não tem relacionamento significativo com **impacto**, mas tem relacionamento fraco e significativo com as três medidas de **reação**. Os níveis de **reações** e **impacto**, estes sim, estão correlacionados forte e significativamente entre si.

Chama a atenção ainda que a aprendizagem é, no presente trabalho, a única variável cuja medida não é de natureza perceptual. As altas correlações entre as demais variáveis critério refletem, em parte, os estilos pessoais dos participantes.

Nesse caso, é preciso tomar cuidado ao extrair conclusões sobre a magnitude desses relacionamentos, antes de examinar a objetividade dessas avaliações. Por outro lado, a medida de aprendizagem utilizada pode ter sido objetiva, mas pouco precisa, se os instrutores que elaboraram os testes não foram capazes de construir itens adequados para medir o alcance dos objetivos instrucionais.

Os dados da literatura a respeito dos relacionamentos entre os níveis de avaliação são contraditórios e inconclusivos. Enquanto Roullier e Goldstein (1993) encontraram um relacionamento positivo entre aprendizagem e transferência, Noe e Schmitt (1986) e Tracey, Tannenbaum e Kavanagh (1995) não encontraram relacionamento significativo entre essas variáveis. No que se refere ao relacionamento do nível de reação com aprendizagem e impacto, Noe e Schmitt (1986) não encontraram relacionamento significativo, enquanto nesta pesquisa, as variáveis referentes ao nível reações estão correlacionadas positivamente com as medidas de aprendizagem e impacto.

Em parte, a força do relacionamento entre essas variáveis de reações com os demais níveis ficou fortalecida devido à alta qualidade psicométrica dos instrumentos de medida, que obtiveram alfas de Cronbach próximos ou superiores a 0,90 (vide Tabelas 1, 2 e 3). Esses ótimos índices deveram-se, provavelmente, ao modo como os itens originais do questionário de Lima, Borges-Andrade e Vieira (1989) foram concebidos. Esses itens, construídos com base no modelo MAIS de Borges-Andrade (1982), incluem múltiplas variáveis (apoio, procedimentos, processos, disseminação e resultados), que propiciam ao treinando uma base abrangente, segura e objetiva para a avaliação dos cursos.

A estrutura empírica dos fatores de **reação**, descrita nas Tabela 1 e 2, com duas subescalas (Fator 1: **reação à programação e ao apoio** e Fator 2: **reações aos resultados, aplicabilidade e expectativas de suporte**), contém menos componentes que os propostos pelo modelo MAIS. As respostas dos participantes ao questionário reuniram em um só fator a satisfação com os resultados imediatos do treinamento, sua aplicabilidade e as expectativas de obter suporte à transferência. Do mesmo modo, os itens de programação e apoio ficaram agrupados em um só fator.

Esta pesquisa de validação mostrou que o nível de reação não é essencialmente unidimensional, podendo ser composto por dois fatores correlacionados entre si que medem aspectos diferentes e igualmente relevantes dos cursos. Novos instrumentos, ao serem construídos, não deveriam conter medidas **globalizantes** e excessivamente sintéticas. Instrumentos mais analíticos e de conteúdo representativo do conceito de reações talvez sejam mais úteis, porque possibilitam retroalimentações mais precisas e relevantes, facilitando, desse modo, o aperfeiçoamento dos sistemas instrucionais desenvolvidos para os treinamentos.

A escala que mede a **satisfação do participante** com o **desempenho do instrutor** também mostrou um ótimo índice de consistência interna (alfa de Cronbach = 0,96). Os resultados dos procedimentos de validação mostraram que essa escala, de natureza primordialmente unidimensional, também pode ser utilizada com duas subescalas: Fator 1, desempenho didático, e Fator 2, entrosamento com os treinandos e domínio do conteúdo, ambas igualmente confiáveis; porém mais estudos deveriam ser feitos para examinar de que modo as variáveis relativas aos cursos e à clientela estão relacionadas com as reações dos treinandos a cada um dos dois fatores de desempenho do instrutor (desempenho didático X domínio do conteúdo e entrosamento com os treinandos).

As avaliações dos participantes relacionadas à **competência interpessoal** do instrutor, no estudo de Abbad (1999), não entraram em nenhum modelo de regressão múltipla *stepwise* como variáveis preditoras de impacto do treinamento no trabalho. A variabilidade das respostas sobre **entrosamento com os treinandos** foi pequena, se comparada à dos demais itens. Além disso, há sinais de **leniência** nos julgamentos dos treinandos sobre esses quesitos. Apesar de integrarem uma escala confiável, as variáveis de reação ao **entrosamento** não parecem influenciar os resultados do treinamento, provavelmente porque se referem a características pessoais do instrutor sobre as quais os treinandos não se sentem à vontade para emitir juízos mais precisos.

Cada elo da corrente de eventos **reações → aprendizagem → impacto no trabalho** parece depender de conjuntos diferentes de variáveis. A aprendizagem parece não depender apenas das características do treinamento. Esse nível pode estar relacionado também com características pessoais da clientela e com o suporte organizacional. As reações, por sua vez, podem depender de combinações diferentes do mesmo tipo de variáveis preditoras de aprendizagem. O impacto do treinamento no trabalho, por outro lado, pode depender fortemente do suporte à transferência e de variáveis de reações aos resultados do treinamento. Todos os níveis de avaliação, ou variáveis critério, parecem depender de certas características da clientela e quase todas de percepções de suporte organizacional. Resultados deste tipo foram encontrados no trabalho de Abbad (1999) e apóiam a idéia de Roe (1997) de que cada elo da cadeia de níveis de avaliação é explicado por diferentes variáveis. As baixas correlações encontradas entre reações e aprendizagem, por exemplo, talvez possam ser mais bem explicadas avaliando-se a natureza e a magnitude dos relacionamentos entre essas variáveis critério e as respectivas variáveis preditoras. A explicação de cada nível dependeria, neste contexto, não apenas do nível anterior de avaliação, mas de uma complexa gama de variáveis pertencentes à organização, ao treinamento, à clientela e, provavelmente, ao ambiente extra-organizacional.

Além disso, parecem muito complexos os relacionamentos entre **aprendizagem, reações e impacto**. Essa complexidade será mais bem compreendida, quando estiverem mais bem explicadas as relações de cada um desses níveis, com as variáveis de suporte organizacional e à transferência, características da clientela e do curso.

A generalidade dos resultados constantes deste trabalho é limitada pelas características da amostra, composta por servidores públicos de alto nível de escolaridade, todos treinados pelo mesmo órgão. Mais pesquisas sobre avaliação de treinamento deveriam ser realizadas em outros ambientes organizacionais, com diferentes treinamentos e amostras de participantes, de modo a ampliar a generalidade dos resultados, bem como testar modelos de avaliação abrangentes que integrem pelo menos os três primeiros níveis de avaliação.

REFERÊNCIAS BIBLIOGRÁFICAS

ABBAD, G. A.

Um modelo integrado de avaliação do impacto do treinamento no trabalho - IMPACT. Brasília, 1999. Tese (Doutorado) – Instituto de Psicologia, Universidade de Brasília.

ABBAD, G. A.;

LIMA, G. B. C.;

VEIGA, M. R. M.

Implantação de um sistema de avaliação de desempenho : métodos e estratégias. **Revista de Administração**, v. 31, n. 3, p. 38-52, 1996.

ABBAD, G. A.;

SALLORENZO, L. H.;

GAMA, A. L. G.

Reação de treinandos a cursos de curta duração : adaptação e vali-

dação de um instrumento. In: XXVIII REUNIÃO ANUAL DE PSICOLOGIA (1998 : Ribeirão Preto). **Anais...** Ribeirão Preto : Sociedade Brasileira de Psicologia, 1998. p. 204. resumo.

ALLIGER, G. M.;

JANAK, E. A.

Kirpatrick's levels of training criteria : thirty years later. **Personnel Psychology**, n. 42, p. 331-342, 1989.

ALVES, A. R.;

PASQUALI, L.;

PEREIRA, M. A. M.

Escala de satisfação com o treinamento - ESAST/Telebrás/UnB. **Revista de Administração de Empresas**, v. 39, n. 1, p. 25-30, 1999.

ALVES, A. R.;

TAMAYO, A.

Sistema de avaliação da Telebrás – SAT. **Revista de Administração**, n. 28, v. 4, p. 73-80, 1993.

BORGES-ANDRADE, J. E.

Avaliação somativa de sistemas instrucionais : integração de três propostas. **Tecnologia Educacional**, v. 11, n. 46, p. 29-39, 1982.

BORGES-ANDRADE, J. E.;

ABBAD, G.

Treinamento no Brasil : reflexões sobre suas pesquisas. **Revista de Administração**, v. 31, n. 2, p. 112-125, 1996.

BORGES-ANDRADE, J. E.;

LIMA, S. M. V.;

REIS, M. J. D.

Formação de pesquisadores : condições de ensino na pós-graduação e variáveis relacionadas. In: INEP-MEC (Org.). **Bolsistas no exterior**. Brasília : INEP, 1989. p. 113-115.

BROAD, M. L.

Management actions to support transfer of training. **Training and Development Journal**, v. 36, n. 5, p. 124-130, 1982.

DRISKELL, J. E.;

WILLIS, R. P.;

COOPER, C.

Effect of overlearning on retention. **Journal of Applied Psychology**, v. 77, n. 5, p. 615-622, 1992.

GAGNÉ, R. M.

Essentials of learning for instruction. 2. ed. Englewood Cliff, NJ : Prentice-Hall, 1968.

GOLDSTEIN, I. L.

Training in work organizations. In: DUNNETTE; HOUGH (Orgs.). **Handbook of industrial and organizational psychology**. 2. ed. California : Consulting Psychology Press, 1991. p. 507-619.

HAMBLIN, A. C.

Avaliação e controle de treinamento. São Paulo : McGraw-Hill do Brasil, 1978.

KIRPATRICK, D. L.

Evaluation of training. In: CRAIG, R. L. (Org.). **Training and development handbook**. 2. ed. New York : McGraw-Hill, 1976. p. 118.1-118.27.

Evaluating training programs evidence vs. proof. **Training and Development Journal**, v. 31, n. 11, p. 09-12, 1977.

LIMA, S. M. V.

Brasil ou exterior? Onde treinar nossos futuros cientistas. In: INEP-MEC (Org.). **Bolsistas no exterior**. Brasília : INEP, 1989. p. 91-111.

LIMA, S. M. V.;

BORGES-ANDRADE, J. E.

Meta-análise de avaliação de treinamento : investigação das relações entre insumo, planejamento, apoio, processo e resultados. **Série Estudos e Pesquisas**, n. 30, 1984.

Meta-análise de avaliação de treinamento. **Revista de Administração**, v. 20, n. 3, p. 39-52, 1985.

LIMA, S. M. V.;

BORGES-ANDRADE, J. E.;

VIEIRA, S. B. A.

Cursos de curta duração e desempenho em instituições de pesquisa agrícola. **Revista de Administração**, v. 24, n. 2, p. 36-46, 1989.

NOE, R. A.;

SCHMITT, N.

The influence of trainee attitudes on training effectiveness : test of a model. **Personnel Psychology**, v. 39, p. 497-523, 1986.

PAULA, S. M. A.

Variáveis preditoras de impacto de treinamento no trabalho : análise da percepção dos treinandos de duas organizações. Brasília, 1992. Dissertação (Mestrado) – Instituto de Psicologia, Universidade de Brasília.

QUIRINO, T. R.;

BORGES-ANDRADE, J. E.;

PEREIRA, W. C. A.

Recursos humanos, conhecimento e tecnologia : avaliação do programa de pós-graduação da Embrapa no Brasil e sugestões de melhorias. Brasília : Embrapa-DRH, 1980. Documentos 2.

REISER, R. A.;

DICK, W.

Instructional planning : a guide for teachers. Needham Heights : Mass, Allyn & Bacon, 1996.

ROE, R. A.

Assumptions and dilemmas in training. **Applied Psychology**, v. 80, n. 2, p. 360-365, 1997.

ROULLIER, J. Z.;

GOLDSTEIN, I. L.

The relationship between organizational transfer climate and positive transfer of training. **Human Resource Development Quarterly**, v. 4, n. 4, p. 377-390, 1993.

STUFFLEBEAM, D. L.

The CIPP model for program evaluation. In: MADAUS, G. F.; SCRIVEN, M.; STUFFLEBEAM, D. L. (Eds.). **Evaluation models** : viewpoints on educational and human services evaluation. Boston : Kluwer, 1984.

TABACHNICK, B. G.;

FIDELL, L. S.

Using multivariate statistics.
New York : Harper & Row, 1996.

TANNENBAUM, S. I.;

YUKL, G.

Training and development in

work organizations. **Annual Review of Psychology**, v. 43, p. 399-441, 1992.

TRACEY, J. B.;

TANNENBAUM, S. I.;

KAVANAGH, M. J.

Applying trained skills on the job : the importance of the work environment. **Journal of Applied Psychology**, v. 80, n. 2, p. 239-252, 1995.

VANLEHN, K.

Cognitive skill acquisition. **Annual Review of Psychology**, v. 47, p. 513-539, 1996.