

*Sistemas de avaliação de desempenho: uma revisão de literatura**

Ione Milani, PhD
PhD Instituto de Psicologia
Universidade de Brasília

Resumo

Este estudo mostrou que as técnicas atuais de avaliação de desempenho são questionáveis quanto aos usos pretendidos. Há problemas relacionados à exatidão, viabilidade, confiabilidade, validade de conteúdo e discriminação. Nenhuma técnica é aplicável para toda avaliação. O estudo inclui sugestões para o desenvolvimento de novas abordagens para avaliação de desempenho.

Palavras-chaves

- sistemas de avaliação de desempenho.
- revisão crítica das técnicas de avaliação de desempenho.
- escalas de comportamentos observáveis.
- escalas de comportamentos esperados.
- administração por objetivos.

* Este trabalho foi desenvolvido pela autora como parte dos requisitos para o grau de doutor em Psicologia, na Western Michigan University, em Kalamazoo, Michigan, EUA, em dezembro de 1983. A autora agradece Norman Peterson, PhD pela orientação durante a elaboração do mesmo.

INTRODUÇÃO

O uso, com múltiplos objetivos, de Sistemas de Avaliação de Desempenho (SAD) é muito conhecido e problemático. SAD são usados para objetivos administrativos, tais como promoção e seleção, bem como para mudança de comportamento dos empregados.

Os objetivos desta revisão de literatura focalizando os SAD são:

- Analisar e questionar a validade dos instrumentos utilizados neste processo;
- Analisar a coerência entre a escolha do instrumento e o objetivo a que se propõe;
- Analisar as próprias razões para a aplicação dos SAD;
- Sugerir um instrumento elaborado a partir dos objetivos do processo de avaliação de desempenho, bem como sugerir um SAD para a organização como um todo e não apenas para as pessoas que fazem parte da mesma.

A primeira seção deste trabalho apresenta as razões mais comuns para aplicação de SAD e a segunda parte trata dos problemas relacionados com as razões citadas para aplicação de tais sistemas.

APLICAÇÕES COMUNS DE SAD

Salário e Compensação: Na área de salário e compensação, as principais razões para o uso de SAD têm sido determinar níveis de pagamento e estrutura de salário (Bass & Barrett, 1981; Kahalas, 1980; McCormick & Tiffin, 1965; Sikula & Sikula, 1978; Smith & Browver, 1977), para determinar ajustamentos ou aumentos de salário (Lacho, Stearns & Villere, 1979; McGregor, 1957), prover justificação escrita para mudança de salário (Meyer, Kay & French, 1965), fazer pagamentos e salários contingentes ao desempenho (Wexley & Yukl, 1977), tomar decisões equilibradas sobre compensações (Lacho et al., 1979; Lefton & Buzzotta, 1978; Locher & Teel, 1977), e determinar salários e incentivos (bônus) justos baseados em méritos ou resultados (Lazer & Wikstrom, 1977).

Decisões Administrativas: Sistemas de avaliação de pessoal têm sido usados também para estabelecer pesquisa e referência básica para decisões sobre pessoal (Oberg, 1972), facilitar tomada de decisão durante os diversos estágios do processo de contratação (Sikula & Sikula, 1978), providenciar informações necessárias para as atividades de planejamento administrativo (Lacho et al., 1979; Lazer & Wikstrom, 1977; Locher & Teel, 1977; Sikula & Sikula, 1978; Smith & Browver, 1977), e avaliar as necessidades de recursos humanos e planejar a realocação (Lazer & Wikstrom, 1977). Portanto, eles têm providenciado informação sobre quando e onde há maior necessidade de supervisão (Jacobs, Kafry & Zedeck, 1980), dados para os administradores utilizarem quando necessário para futuras seleções (Lewinson, 1976), e dados para julgar práticas de emprego e política empresarial relacionadas à saúde e segurança do trabalho bem como a serviços e benefícios (Sikula & Sikula, 1978).

Dados de avaliação têm sido também usados para decisões relacionadas a demissões e transferências (Bass & Barrett, 1981; Lacho et al., 1979; Locher & Teel, 1977; McCormick & Tiffin, 1965; McGregor, 1957; Oberg, 1972; Sikula & Sikula, 1978; Wexley & Yukl, 1977); para ações disciplinares tais como, suspensão, retirada ou um decréscimo de salário, ou demissão (Jacobs et al., 1980); e decisões administrativas, tais como, transferências e alocação de recursos financeiros (DeCotiis & Petit, 1978). Além disso, tais dados têm sido também utilizados para decisões sobre promoções (Bass & Barrett, 1981; Jacobs et al., 1980; Lacho et al., 1979; McGregor, 1957; Sikula & Sikula, 1978; Smith & Browver, 1977), identificar pessoas com potencial para promoção (Lazer & Wikstrom, 1977; Lefton & Buzzotta, 1978; Locher & Teel, 1977; Oberg, 1972; Wexley & Yukl, 1977), ajudar a tomada de decisões quanto ao grau de promoção (Krausz, 1980), descobrir pessoas com alto potencial para ocupar cargos administrativos (Smith & Browver, 1977) e fazer estimativa do potencial dos subordinados para o desenvolvimento e promoção (Lazer & Wikstrom, 1977).

Aumento da Produtividade: Dados provindos dos sistemas de avaliação têm sido utilizados para estabelecer objetivos (Lacho et al., 1979; Locher & Teel, 1977; Wexley & Yukl, 1977), como uma ajuda para criar e manter um nível satisfatório de produção (Sikula & Sikula, 1978), como uma base para modificar ou mudar comportamentos direcionando hábitos de trabalho mais efetivos (Lewinson, 1976) e para tornar os executores mais produtivos (Lazer & Wikstrom, 1977).

Além disso, sistemas de avaliação têm sido utilizados como um meio de aumentar a motivação dos subordinados a fim de melhorar o desempenho no trabalho dando-lhes *feedback* sobre o que se espera deles, como a tarefa está sendo bem executada e como eles podem desempenhar melhor suas funções (Bass & Barrett, 1981; Jacobs et al., 1980; Lacho et al., 1979; Lazer & Wikstrom, 1977; Lefton & Buzzotta, 1978; Lewinson, 1976; Locher & Teel, 1977; Meyer et al., 1965; Oberg, 1972; Sanders & Peay, 1974; Smith & Browver, 1977; Wexley & Yukl, 1977), ou sugerindo mudanças de comportamento (McGregor, 1957) ou satisfazendo a "necessidade de conhecer" (Haynes, 1978).

Treinamento e Desenvolvimento: Dados de avaliação de pessoal têm sido também usados para medidas de desempenho (Lazer & Wikstrom, 1977), aconselhamento e orientação de executores (Bass & Barrett, 1981; McGregor, 1957), ajudar supervisores a observar mais de perto seus subordinados e treiná-los melhor (Oberg, 1972) e melhorar a comunicação entre supervisores e subordinados (Lacho et al., 1979).

Em adição, SAD têm sido utilizados para avaliar necessidades de treinamento (Bass & Barrett, 1981; DeCotiis & Petit, 1978; Krausz, 1980; Lacho et al., 1979; Locher & Teel, 1977; Sikula & Sikula, 1978; Wexley & Yukl, 1977), avaliar programas de treinamento (Jacobs et al., 1979; Wexley & Yukl, 1977) e determinar necessidades de desenvolvimento gerencial (Lacho et al., 1979; Lazer & Wikstrom, 1977). Dados de avaliação têm também ajudado a identificar oportunidade para desenvolvimento e crescimento na carreira (DeCotiis & Petit, 1978; Haynes, 1978; Lacho et al., 1979), levando o administrador a um interesse nas necessidades individuais ▶

de treino e desenvolvimento (Henderson, 1976; Wexley & Yukl, 1977).

Critérios de Medidas: O processo de avaliação de desempenho tem ajudado nos procedimentos de validação de seleção e promoção (Lacho et al., 1979); bem como de tomada de decisões referentes a recrutamento, seleção, alocação, instrução (Sikula & Sikula, 1978) e estabilidade (Krausz, 1980). As técnicas de avaliação de desempenho deveriam prover critérios para validação de estudos em seleção e pesquisas sobre pessoal (DeCotiis, 1978; Jacobs et al., 1980; Locher & Teel, 1977; McCormick & Tiffin, 1965; Sikula & Sikula, 1978), bem como providenciar documentação para decisões de ações positivas, diminuindo assim a probabilidade de custos processuais (Lacho et al., 1979; Lefton & Buzzotta, 1978; Locher & Teel, 1977).

Planejamento de Recursos Humanos: Finalmente, um processo de avaliação pode ser utilizado para revelar deficiências na administração, indicar as habilidades mais necessárias para uma organização (Oberg, 1972), providenciar base para avaliação da eficiência produtiva da organização como um todo bem como das unidades que a compõem (Wexley & Yukl, 1977), manter o controle da organização (Cummings e Schwab, 1973), atingir ótimos resultados, para que ela alcance seus objetivos (Henderson, 1976; Lefton & Buzzotta, 1978), medir e melhorar a eficiência na utilização dos recursos humanos da organização (Cummings e Schwab, 1973) e avaliar a efetividade dos horários e métodos de trabalho, a estrutura organizacional, estilos gerenciais, condições de trabalho e equipamento (Wexley & Yukl, 1977).

PROBLEMAS DOS SISTEMAS DE AVALIAÇÃO DE DESEMPENHO RELACIONADOS COM OS OBJETIVOS

A utilização dos SAD para muitos objetivos cria conflitos entre os mesmos (Beer, Ruh, Dawson, McCaa & Kavanaugh, 1978; Kahalas, 1980; Lazer e Wikstrom, 1977; Meyer et al., 1965; Williams, 1972). Este conflito, às vezes, leva o sistema de avaliação a se tornar apenas uma discussão sobre salário durante a qual os administradores justificam suas ações (Meyer et al., 1965). Dois objetivos opostos, tais como determinar aumento anual de salário e motivar pessoas para melhorar o desempenho, requereu comportamentos gerenciais como juiz e conselheiro, simultaneamente (Kahalas, 1980). Assumindo que um gerente deva decidir sobre o corte de salário a um subordinado, qualquer tentativa para motivá-lo pode ser inefetiva; isto é, decisões sobre pagamento podem interferir com um bom aconselhamento (Lawler, 1981). Colocar os dois objetivos em um só programa não somente produz conflito como também pode levar a pior desempenho no futuro (Williams, 1972) ou pode haver pequena ou nenhuma influência futura no desempenho da tarefa (Meyer et al., 1965).

Smith & Browver (1977) afirmaram que os objetivos da avaliação de desempenho são diferentes dos objetivos de desenvolvimento. Por exemplo, objetivos administrativos e comportamentais são distintos e assim devem ser tratados. Objetivos gerenciais são os usos que a organização faz da avaliação de desempenho para os

empregados, tais como para promoção, seleção, treinamento, disciplina etc, enquanto que objetivos comportamentais são relacionados aos esforços próprios dos indivíduos para melhorar seus comportamentos no trabalho (Ralph, 1980).

Williams (1972) sugeriu que se distinguísse pelo menos três formas diferentes de avaliação: para o desempenho atual, para o potencial e para a determinação de salário. Yager (1981) separou os objetivos para avaliação de desempenho daqueles utilizados para revisão de desempenho: por exemplo, a motivação do executor através do *feedback*, o qual deveria ocorrer regular e freqüentemente entre supervisores e subordinados, é um objetivo do sistema de avaliação de desempenho; aumento de salário, contudo, é um objetivo da revisão de desempenho. Lawler (1981) sugeriu que a solução para os conflitos entre os objetivos poderia ser sessões separadas para cada objetivo.

Locher & Teel (1977) recomendaram que se desse mais ênfase na definição dos objetivos da avaliação do que nas técnicas. Ralph (1980) notou que os objetivos deveriam ser definidos antes da construção dos instrumentos de avaliação, mas freqüentemente isto não acontece. Uma recente revisão de estudos relacionados ao uso das escalas *Behaviorally Anchored Rating Scales (BARS)* revelou que nenhum dos procedimentos descritos naqueles estudos indicavam ao avaliador, ou mesmo para o leitor, o objetivo da avaliação (Jacobs et al., 1980). Yager (1981) afirmou que a falta de objetivos claros explicava por que muitos sistemas de avaliação são fracamente planejados. Determinar o objetivo específico a ser atingido pelo programa de avaliação deveria ser o primeiro passo no planejamento de tais sistemas (Locher & Teel, 1977; Ralph, 1980), por ser este o passo mais crítico (Sanders & Peay, 1974).

De acordo com Lazer & Wikstrom (1977) a avaliação de um sistema deveria envolver um exame sobre até que ponto o sistema atinge seus objetivos. Jacobs et al., (1980) sugeriu que a escolha da técnica a ser usada em um processo de avaliação de desempenho deveria ser baseada no objetivo da avaliação. Além disso, os objetivos da avaliação deveriam ser enfatizados para os avaliadores. Cada objetivo requer metodologias diferentes para coletar as informações. Em outras palavras, cada objetivo pode requerer diferentes técnicas de avaliação. Além do mais, requisitos legais exigem que os sistemas de avaliação deveriam ser racionais, incluindo objetivos específicos e técnicas para atingir estes objetivos (Kahalas, 1980).

Oberg (1972) concluiu que a falta de efetividade nos SAD resulta da falha em casar (*match*) a técnica selecionada com o objetivo estabelecido para a avaliação. Em outras palavras, não é dada muita atenção sobre a escolha de uma técnica específica que seja a melhor para atingir determinado objetivo de avaliação. DeCotiis & Petit (1978) afirmam que, apesar da existência de uma grande variedade de instrumentos de avaliação de desempenho, apenas alguns estudos investigaram se eram adequados para o objetivo de avaliação. Em estudo recente (1977), DeCotiis sugeriu que o objetivo final do instrumento deveria ser considerado quando um SAD está sendo desenvolvido. Finalmente, mesmo quando os objetivos são identificados, outro problema nos atuais ►

SAD é a discrepância entre os objetivos identificados e os usos do instrumento (Lazer & Wikstrom, 1977).

Em resumo, SAD têm sido utilizados para atingir muitos e os mais variados objetivos. O uso de um SAD para atingir diferentes objetivos simultaneamente pode ser problemático, porque cada objetivo requer estratégias diferentes e dados específicos do desempenho individual. A seguir, far-se-á análise dos diferentes dados necessários para os vários usos dos SAD.

DADOS NECESSÁRIOS PARA CADA OBJETIVO ESPECÍFICO

Supõe-se que as técnicas utilizadas em sistemas de avaliação forneçam informações sobre diferentes aspectos do desempenho do executor. Assim, os dados a serem coletados deveriam ser adequados para o objetivo final de cada avaliação, porque objetivos diferentes podem requerer dados diferentes. Portanto, planejadores e avaliadores de SAD deveriam conhecer quais dados são necessários para atingir os objetivos propostos. Há dois tipos importantes de dados: os resultados do desempenho ou as medidas dos comportamentos que se supõe produzem aqueles resultados. Neste item, as informações exigidas para objetivos diferentes são examinadas; o primeiro conjunto de objetivos fundamentalmente exige medidas de resultados e o segundo conjunto requer medidas de comportamentos.

Decisões administrativas: A maioria das decisões gerenciais e muitas atividades são planejadas com a finalidade de atingir os objetivos da organização através do melhor uso dos recursos humanos. Os dados coletados no processo de avaliação de desempenho podem ajudar o gerente a tomar decisões relacionadas à contratação, demissão, transferência ou promoção. Um gerente necessita de dados corretos para tomar decisões razoáveis relacionadas às situações de trabalho dos subordinados. Muitos problemas podem ser prevenidos quando se tem dados válidos e confiáveis como base para decisões sobre pessoal. Para o objetivo de avaliação de desempenho, neste caso, os dados que seriam válidos e necessários poderiam ser os resultados esperados e medidos para tarefa. Ou seja, os resultados produzidos pelo executor.

A validação dos procedimentos de seleção requer uma relação significativa entre os prognósticos *predictors* e os critérios (*American Psychological Association*, 1980). O desempenho deveria ser avaliado com medidas de critério objetivas, focalizando nos resultados esperados e produzidos pelo executor; assim, os requisitos "predictors" utilizados para selecionar as pessoas para uma dada tarefa serão relacionados de maneira mais significativa às medidas atuais do desempenho na tarefa.

Dar *feedback* ao subordinado requer conhecimento preciso de como a tarefa está sendo bem executada. *Feedback* deveria ser preciso, baseado em áreas específicas do desempenho e deveria representar medidas de resultados relevantes (Rummler, 1972). A informação dada em uma intervenção *feedback* refere-se à quantidade, qualidade ou natureza dos resultados do desempenho (Nadler, Mirvis & Camman, 1976). As intervenções *feedback* são mais efetivas quando elas são mais fre-

quentes, relacionadas aos ciclos temporais de uma tarefa (Rummler, 1972), baseado na suposição de que a frequência na apresentação de *feedback* aumenta a exatidão e o momento oportuno da informação recebida.

Planejamento de Recursos Humanos: A alocação de recursos humanos é determinada pela lotação existente e a demanda por pessoas para as tarefas nos vários níveis da organização. Para planejar a melhor alocação de recursos humanos é necessário conhecer o que leva à lotação atual existente, bem como a demanda por pessoas. Uma fonte de informação são os dados coletados em um sistema de avaliação onde o nível do desempenho é relacionado a que resultados esperados foram produzidos. Esta informação facilitaria o planejamento gerencial. Por exemplo, a demanda por mais pessoas pode, às vezes, ser atendida através da melhora na produtividade de um pequeno número de executores; outras vezes, contudo, são necessários mais executores. Aqui também os dados obtidos da observação dos resultados esperados e atingidos pelo executor devem ser confiáveis e válidos para o objetivo de planejamento gerencial.

A organização pode ser melhorada pela identificação de pessoas com potencial para promoção e através do conhecimento das necessidades de desenvolvimento (Oberg, 1972). O processo de avaliação de desempenho pode ser utilizado para indicar as habilidades mais necessárias à organização (Jacobs et al., 1980), mostrando o que um executor deve realizar em uma dada tarefa ou coletando dados que refletem as deficiências do executor. Um modelo no qual os requisitos para uma tarefa são descritos em termos de realizações esperadas a serem atingidas e as habilidades necessárias para a posição vaga ajudaria a selecionar ou promover a pessoa certa — um executor habilidoso que preencheria as necessidades da organização.

Treinamento e Desenvolvimento: O objetivo de um sistema de treinamento é preparar pessoas para executar suas tarefas no padrão exigido, através da mudança de comportamentos. Assim, os comportamentos exigidos devem ser determinados e a mudança deve ser medida em relação a padrões específicos bem estabelecidos (Warren, 1979). Gilbert (1967) sugeriu algumas regras a se seguir no levantamento de necessidades de treinamento. Uma regra está resumida na fórmula $D = M - I$, onde deficiência (D) é igual ao domínio *mastery* (M) ou a lista de comportamentos que se assume sejam necessários para o domínio de um assunto menos (—) o repertório inicial (I) ou os comportamentos que o treinando já apresenta quando o treinamento se inicia. Portanto, a deficiência do executor no desempenho atual seriam os dados necessários para se estabelecer objetivos para o treinamento, o que significa a diferença entre os padrões de desempenho estabelecidos (*prescribed mastery*) e o repertório atual do executor. Outra regra de Gilbert para levantar necessidades de treinamento se refere à necessidade de distinguir entre as habilidades (*acquisition*) do executor e as realizações esperadas. Habilidades se referem ao que deve ser aprendido ou o resultado produzido pela aprendizagem, enquanto que as realizações esperadas se referem ao valor dado ao que foi aprendido. As deficiências nas habilidades fazem diferenças nas realizações esperadas, mesmo uma pequena mudança nas habilidades pode produzir grandes mu- ▶

danças nas realizações esperadas. Para o objetivo de levantar necessidades de treinamento, as habilidades devem ser medidas, não as realizações esperadas. Portanto, os dados a serem coletados para se estabelecer as necessidades de treinamento se resumem em uma lista das deficiências em termos de comportamentos/habilidades.

Um processo de avaliação deveria indicar as deficiências do executor em termos das diferenças entre os padrões estabelecidos e o desempenho atual. Assim fazendo, as necessidades de treinamento são levantadas e a organização pode melhorar pela melhora do desempenho através de um programa de treinamento (Jacobs et al.).

Sistemas de Pagamento: Os sistemas de pagamento são um motivador potencial para melhorar o desempenho dos executores, o qual deveria ser um critério em que está baseado o sistema de pagamento; o critério para avaliar o desempenho do executor deveria ser objetivo e incluir todos os comportamentos para uma tarefa (Lawler, 1981). Uma razão para aumentar salários baseada no mérito é que as empresas podem perder executores habilidosos que procuram níveis de salários mais elevados em outras empresas (Henrici, 1980). Assim, o bom executor deve receber mais que o mau executor, para que seja mantido na organização (Lawler, 1981). Lawler mostrou que as taxas de *turnover* são mais baixas quando a satisfação em termos de salários é maior.

De acordo com Gilbert (1978), prestar atenção somente a comportamentos leva a maior ênfase na quantidade de energia que as pessoas dispendem em vez da qualidade do que foi realizado. Assim, o comportamento é visto como um fim ao invés de um meio para se atingir um fim (isto é, a realização esperada, que é a medida real da competência). O desempenho deve ser definido cuidadosamente. Os comportamentos são apenas parte do desempenho e não devem ser confundidos com ele. Observar e medir os comportamentos não indica necessariamente qual é o desempenho atual. A transação total que inclui comportamento e o que é realizado ou produzido pelo mesmo é o que se chama desempenho (Gilbert, 1978). O mesmo autor define desempenho valioso *worthy* como o desempenho no qual o valor das realizações esperadas excede o custo do comportamento. Portanto, o desempenho pode ser melhorado aumentando o valor das realizações e diminuindo a energia dispendida na sua execução. Assim, o incentivo das pessoas deve ser conectado às realizações esperadas, as quais devem ser medidas e avaliadas através das dimensões de qualidade, produtividade e custo.

Os dados coletados para o objetivo de aumento de salário deveriam ser precisos; isto é, as medidas identificadas devem refletir o verdadeiro nível do desempenho do executor. De acordo com Jacobs et al. (1980), exatidão é determinada pela exclusão de erros de escalas, tais como halo, tolerância ou leniência, tendência central, contraste e assim por diante, e a exatidão é diminuída pela falta de especificação dos verdadeiros níveis de desempenho. Erros de escalas podem levar a ações salariais injustas, tais como negar aumento de salário a um bom executor. Outrossim, reações emocionais podem resultar de comparações entre empregados que trabalham lado a lado e são tratados injustamente. Quando

os dados não são corretos, as conseqüências para o desempenho podem produzir maus efeitos, tais como comportamentos supersticiosos emitidos quando o reforçamento accidental foi contingente a um dado desempenho. Bons executores podem se afastar do emprego se forem injustamente não recompensados pelo seu desempenho, ou uma reação emocional pode ser derivada de uma conseqüência injusta, onde o mau executor é recompensado da mesma maneira que o bom executor.

TÉCNICAS ATUALMENTE UTILIZADAS PARA AVALIAÇÃO DE DESEMPENHO

Behaviorally Anchored Rating Scales: As escalas chamadas *Behaviorally Anchored Rating Scales (BARS)* ou *Behavioral Expectation Scales (BES)* são dois termos que se referem ao mesmo instrumento de avaliação baseado nos comportamentos (Latham & Wexley, 1981). Este método contém a Técnica do Incidente Crítico (CIT, *Critical Incident Technique*), a qual envolve uma lista de incidentes críticos de comportamentos efetivos e inefetivos no trabalho observado nos últimos 6 a 12 meses. Os observadores devem conhecer os objetivos de uma dada tarefa e ver freqüentemente as pessoas que a executam. Entre os observadores podem ser incluídos os supervisores, colegas, subordinados e clientes. Através de entrevista com os observadores, os incidentes dos comportamentos efetivos e inefetivos são listados, a fim de formular os requisitos críticos da tarefa (Latham & Wexley, 1981). Os executores não são entrevistados porque eles não se sentiriam à vontade para descrever os comportamentos inefetivos (Vroom & Maier, 1961). Latham & Wexley (1981) recomendam que pelo menos 30 pessoas deveriam ser entrevistadas para um total de aproximadamente 300 incidentes. Os requisitos críticos da tarefa são formulados de uma lista de incidentes observados, mas freqüentemente muitos itens da lista original dos incidentes observados são descartados na composição dos requisitos críticos da tarefa. Alguns incidentes são descartados porque se sobrepõem a outros, e alguns são descartados porque as pessoas que elaboram as BARS não concordam que eles deveriam ser retidos (Schwab, Heneman & DeCotiis, 1975). Este processo de formulação dos requisitos críticos da tarefa baseado na lista de incidentes observados é chamado de reformulação (Latham & Wexley, 1981).

De acordo com Flanagan (1954), supõe-se que a CIT seja uma coleção de observações diretas do comportamento que facilita inferências e predições sobre o desempenho do executor. Para incluir todos os comportamentos críticos, o número de incidentes variam de acordo com a tarefa. Um cargo de supervisão precisaria coletar entre 2.000 a 4.000 incidentes; "tarefas que exigem muita habilidade" exigiram entre 1.000 a 2.000 incidentes definidos como comportamentos extremos, sendo efetivos ou inefetivos para atingir os objetivos da atividade (Flanagan, 1954).

Os requisitos críticos da tarefa são então categorizados, e estas categorias se tornam os critérios para avaliar o desempenho dos executores. Os itens de cada categoria colocados *rated* numa escala de 7 (sete) pontos onde 7 (sete) representa desempenho excelente, 4 (quatro) ►

representa a média e 1 (um) representa desempenho pobre. O passo final na criação da escala envolve a elaboração dos itens em cada categoria para comportamentos esperados (por exemplo, "O executor datilografa, em média, 60 palavras por minuto" se torna em "espera-se que o executor datilografe 60 palavras por minuto"). Quando a escala é usada para avaliação de desempenho, o avaliador não necessariamente observa o comportamento do executor ao colocar na escala o desempenho do mesmo; ao invés disto, recoleciona, faz estimativas e, com frequência, faz suposições (Latham & Wexley, 1981).

Escalas de Observação Comportamental: As Escalas de Observação comportamental (*BOS Behavioral Observation Scales*) foram desenvolvidas para superar as limitações das BARS, tais como descarte de acidentes críticos gerados na análise de tarefa e a falta de observação direta do desempenho do executor na aplicação destas escalas (Latham, Fay & Saari, 1979). O desenvolvimento das BOS é essencialmente o mesmo das BARS. A diferença é que, nas BOS, a escala de 5 pontos é usada para cada item comportamental onde 0 (zero) significa que o executor emitiu aquele comportamento 0-64% do tempo, 1 (um) para 65-74% do tempo, 2 (dois) para 75-84% do tempo, 3 (três) 85-94% do tempo e 4 (quatro) para 95-100% do tempo (Latham & Wexley, 1981).

Gerência por Objetivos: O sistema chamado Gerência por objetivos (MBO Management by Objectives) é um processo de avaliação de desempenho onde os supervisores e subordinados juntos identificam seus objetivos comuns e definem as áreas mais importantes de responsabilidade individual, em termos de resultados esperados (Odiorne, 1979). No início de cada ano orçamentário, os gerentes e subordinados formulam o desempenho alvo para aquele ano, e no final do mesmo, eles comparam o alvo com os resultados. A condição básica para o sucesso do sistema MBO está na habilidade do gerente monitorar o subordinado de maneira que os objetivos pré-estabelecidos sejam atingidos.

TÉCNICAS UTILIZADAS ATUALMENTE E OS OBJETIVOS DOS SISTEMAS DE AVALIAÇÃO DE DESEMPENHO

Como foi notado antes, diferentes objetivos de SAD requerem dados diferentes. A seguir serão examinados as principais características das técnicas atuais utilizadas para sistemas de avaliação e até onde os dados obtidos com o uso destas técnicas atingem tais requisitos.

BARS: De acordo com Blood (1974), BARS podem ser utilizadas como um instrumento para levantar necessidades de treinamento. A informação coletada na utilização de tais escalas é útil para levantar necessidades de treinamento, porque elas incluem identificação das deficiências do executor e os comportamentos necessários para um bom executor. Os comportamentos esperados listados na elaboração das escalas BARS, contudo, não são necessariamente emitidos pelos executores, o que pode gerar problemas quando os dados coletados são usados para levantar necessidades de treinamento. Podem os dados coletados ser considerados o verdadeiro

nível de desempenho do executor? São eles realmente o reflexo das deficiências do executor? Infelizmente, nenhuma pesquisa tem sido conduzida para investigar o uso das BARS como um instrumento para levantar necessidades de treinamento (Jacobs et al., 1980).

Numa comparação das BARS com um programa orientado para traços, Ivancevich (1980) mostrou que o uso das BARS estimulou níveis altos de desempenho. Entretanto, como não havia grupo de controle no procedimento experimental, não é possível afirmar definitivamente que o sistema de BARS teria causado a melhora, o que poderia ter sido causado pela novidade do sistema. Beatty, Schneir & Beatty (1977) demonstraram melhora no desempenho como resultado da aplicação das BARS.

Analisando os comportamentos do indivíduo no trabalho em várias dimensões, as BARS podem ser utilizadas como base para promoção, mas somente aquelas dimensões comportamentais que sobrepõem as posições presentes e as futuras deveriam ser consideradas (Jacobs et al., 1980). Os mesmos autores notaram que as BARS têm sido utilizadas para a validação de procedimentos de seleção. Entretanto, durante o processo de reformulação *retranslation* na elaboração das BARS, muitos itens foram descartados da lista original de incidentes, o que levou Atkin Colon (1978) a questionar se o domínio da tarefa *job domain* seria mantido em todos os passos da elaboração das BARS.

As BARS podem ser utilizadas, em parte, para outros objetivos, tais como suspensão disciplinar, retirada ou aumento de salário, ou demissão quando há incidentes suficientes para justificar tal avaliação (Jacobs et al., 1980). Entretanto, como os autores mostraram, a utilidade dos dados coletados das BARS para o objetivo de facilitar decisões sobre pessoal é questionável, uma vez que elas focalizam em comportamentos em vez de realizações esperadas, o que leva à conclusão de que o desempenho global não é representado completamente. Esta falta de informação apresenta dificuldades para se distinguir os bons e maus executores, o que pode levar o supervisor a tomar decisões injustas com relação à suspensão disciplinar, retirada ou aumento de salário, e assim por diante. Pela mesma razão, os dados utilizados na elaboração das BARS não parecem adequados para o objetivo de facilitar tomada de decisões relacionadas a aumento de salário e compensação.

BOS: Latham et al. (1979) afirmaram que o sistema chamado BOS leva por si só para programas de treinamento, uma vez que especifica as deficiências do executor em termos de habilidades e conhecimento. Esta afirmação parece razoável tendo em vista que a informação necessária para levantar necessidades de treinamento incluí as deficiências que impedem as realizações esperadas para as várias tarefas. As BOS têm sido utilizadas como um instrumento para avaliar a eficácia em treinar gerentes para o desenvolvimento de habilidades interpessoais (Latham & Wexley, 1977). Além disso, os autores citados afirmaram que as BOS também facilitaram decisões administrativas. Isto ocorreria, contudo, somente quando havia uma informação completa de como o executor desempenhava bem a sua tarefa, e as BOS não fornecem toda informação necessária, uma vez que não medem todas as realizações esperadas. Os dados necessários para atingir o objetivo de facilitar decisões ad-

ministrativas devem conter todas as realizações esperadas e medidas. O mesmo pode ser dito com relação ao uso das BOS para determinar aumento de salário e compensação, para melhorar o desempenho e para diagnóstico e desenvolvimento organizacional.

MBO: Supõe-se que MBO seja útil para atingir muitos objetivos, tais como procedimento de seleção, salário e compensação, melhora do desempenho, e assim por diante. Esta técnica não parece útil para levantar necessidades de treinamento, uma vez que os dados obtidos deste sistema não permitem avaliar as deficiências de comportamento do executor que resultam no desempenho atual. Em outras palavras, os comportamentos que devem ser mudados para atingir os objetivos propostos não aparecem claramente nos dados resultantes do sistema MBO.

Estes dados podem ser úteis para objetivos de *feedback*, porque esta técnica permite ao supervisor e subordinados estabelecerem objetivos para um período de tempo no fim do qual o supervisor compara o nível do desempenho atual com os objetivos previamente estabelecidos. Assim, *feedback* seria baseado em resultados relevantes comparados com um contrato mútuo predefinido para atingir objetivos específicos. A deficiência está no fato de que este *feedback* seria apresentado apenas uma vez por ano, o que o torna menos efetivo. Este *feedback* indicaria apenas a necessidade de mudança e não quais as mudanças a serem feitas.

PROBLEMAS RELACIONADOS COM AS TÉCNICAS UTILIZADAS ATUALMENTE

Os resultados de qualquer processo de avaliação deveriam refletir o verdadeiro nível do desempenho do executor. A metodologia utilizada deveria excluir erros de escalas, tais como halo, leniência, tendência central, lógica, contraste, proximidade e semelhante a mim (Jacobs et al., 1980).

O efeito halo consiste em generalizações inadequadas, feitas pelo avaliador, de um aspecto do desempenho do executor para todos os aspectos do desempenho da pessoa (Latham & Wexley, 1981). Jacobs et al. (1980) demonstraram que resultados de estudos comparando BARS com escalas de pontos somados, escalas gráficas e escalas não ancoradas, não têm diferença com relação ao efeito halo. As escalas de pontos somados consistem de uma escala onde cada dimensão do desempenho é expressa de acordo com um número de afirmações que tornam os componentes distintos tendo um item do tipo Likert com 4 pontos de 0 (zero) representando "concordo fortemente", até 3 (três) representando "discordo fortemente".

Comparando BARS, escalas de pontos somados e os instrumentos baseados em traços, Borman (1979) demonstrou que o instrumento em si não fornece um aumento na exatidão. Em outras palavras, as BARS não oferecem mais exatidão do que outras medidas psicométricas com instrumentos mais simples. Kingstrom Bass (1981) descobriu resultados semelhantes ao rever 51 estudos sobre o uso das BARS: 35 dos estudos indicaram que as BARS têm menor suscetibilidade ao efeito halo, enquanto 14 estudos indicaram maior suscetibilidade e 2 estudos não acharam diferenças entre os instrumentos.

Jacobs et al. (1980) também questionaram o nível absoluto do efeito halo. Alguns autores medem seu nível através do grau de independência das dimensões. As dimensões do desempenho representadas nos incidentes das BARS devem ser medidas independentes, a fim de minimizar a redundância e providenciar uma informação de desempenho única. A intercorrelação das dimensões das BARS tem sido investigada por vários pesquisadores. Revendo estes estudos, Schwab, Heneman & DeCotiis (1975) concluíram que somente três estudos comparando as BARS com outros instrumentos tiveram resultados confusos e portanto não dão suporte à intercorrelação das dimensões das BARS. Os outros estudos que examinaram esta questão não o fizeram em contexto comparativo. A maior dificuldade parece ser a falta de definição operacional padrão para o que seja o efeito halo; como concluíram Kingstrom & Bass (1981), os pesquisadores têm interpretações diferentes deste efeito.

Supõe-se que as BARS são menos suscetíveis ao efeito leniência que as outras técnicas de avaliação do desempenho, devido às dimensões ambíguas desenvolvidas pelo procedimento (Smith & Kendall, 1963). O erro chamado leniência é a tendência do avaliador colocar os indivíduos em níveis mais altos ou mais baixos do que eles realmente merecem, através do uso dos extremos da escala e mostrando uma falta de discriminação da efetividade do avaliado (Jacobs et al., 1980).

Kingstrom & Bass (1981) fizeram 36 comparações separadas entre as BARS e outras escalas com relação ao efeito leniência. Os resultados mostraram que ou não há consistente redução do erro leniência através do uso das BARS, ou as diferenças entre as BARS e outras escalas não eram significantes. Foram feitas algumas tentativas para explicar a inconsistência destes resultados. Zadeck, Kafry & Jacobs (1976) sugeriram que muitas outras variáveis além do tipo das escalas, (tais como o tipo de resposta requerido pelos avaliadores e o *frame* de referência para as respostas) podem levar a interpretações diferentes e, conseqüentemente, a resultados diferentes. A falta de critérios externos, sugerida por DeCotiis (1977), parece ser uma questão importante que deveria ser levada em consideração quando se tenta explicar aquelas inconsistências dos resultados.

A questão principal com relação aos erros das escalas é: quais são as variáveis que afetam as técnicas de avaliação de desempenho, tornando-as suscetíveis a esses erros? Por exemplo, dados coletados de observação direta de comportamentos operacionalmente definidos ou realizações esperadas têm menor probabilidade de serem afetados pelos erros de escalas. A suscetibilidade de inferências no procedimento de laboração e uso das BARS podem explicar sua suscetibilidade ao erro chamado leniência (ver também Borman, 1978).

A correspondência entre o conteúdo da avaliação e o conteúdo da tarefa deve ser alta, para que haja maior probabilidade de exatidão na informação sobre o desempenho (DeCotiis, 1977). As medidas de validade de conteúdo mostram a representatividade e relevância dos itens que compõem os instrumentos em termos de requisitos críticos da tarefa (Latham & Wexley, 1981). A importância deste critério é enfatizada pela Comissão que trata da igual oportunidade de emprego, *Equal Employment Opportunity Commission*, quando afirma que as di- ▶

mensões do desempenho deveriam ser baseadas em uma cuidadosa análise de tarefa. Atkin & Colon (1978) parecem não considerar a Técnica do Incidente Crítico como uma forma de análise de tarefa, quando afirmam que as BARS não são baseadas em análises de tarefa, uma vez que esta raramente precede a elaboração das BARS.

A validade de discriminação pode ser definida como "um esforço para distinguir os executores mais eficientes dos menos eficientes" (Jacobs et al., 1980). Investigadores encontraram resultados confusos entre estudos relacionados com este tipo de validade para as BARS (Jacobs et al., 1980; Kingstrom & Bass, 1981). Borman (1979) encontrou menor validade de discriminação para as BARS, comparadas com escalas de pontos somados e de traços. Campbell, Dunnette, Arvey & Hellervik (1973) encontraram que as BARS têm melhor validade discriminante e convergente. Bernardin, Alvares & Cranny (1976), contudo, questionaram aqueles resultados baseados sobre outros dados coletados no mesmo estudo, tais como concordância entre avaliadores, a qual era maior para as escalas de pontos somados. Validade convergente é o grau de concordância entre dois tipos de medidas. As BARS não podem ser consideradas superiores a outros instrumentos em termos da validade convergentes, uma vez que os resultados de vários estudos são confusos (Jacobs et al., 1980; Kingstrom & Bass, 1981).

Confiabilidade é um critério importante que deveria ser atingido por qualquer sistema de avaliação de desempenho. Uma medida confiável deveria dar uma informação precisa e consistente do desempenho do executor. A consistência das informações e a concordância entre os avaliadores na coleta dos dados provê o nível de confiabilidade para esta medida, a qual mostra uma confiabilidade perfeita quando a correlação é 1.0. Entre avaliadores, a correlação deveria ser pelo menos .60, de acordo com Latham & Wexley (1981).

DeCotiis (1977) afirmou que as BARS têm menor confiabilidade (.64) que as escalas ancoradas numericamente *Numerically Anchored Rating Scales* (.81) e as escalas baseadas em traços (.82). Destes estudos, o mesmo autor concluiu que ambas as escalas numéricas e de traços têm a menos uma confiabilidade satisfatória para os avaliadores. Campbell et al. (1973) e Bernardin et al. (1976) também encontraram maior concordância entre avaliadores nas escalas de pontos somados. Além disso, Bernardin (1977) concluiu que não há maior confiabilidade para os avaliadores e discriminação através dos avaliados para as BARS que para as escalas de pontos somados. A metodologia usada nas BARS resulta em "um modesto nível de concordância entre os avaliadores", concluíram Jacobs et al. (1980), numa revisão de estudos que investigavam a confiabilidade das BARS. E mais, as BARS "não são a panacéia para obter alta confiabilidade para os avaliadores" (Schwab et al., 1975).

Uma das limitações das BARS, apontada por Schwab et al. (1975), refere-se ao número de incidentes perdidos de um passo para outro na elaboração das escalas. Devido ao fato de que muitos incidentes descartados podem representar informações de valor sobre a tarefa, o avaliador pode ter dificuldade em passar os comportamentos observados em dimensões específicas e decidir a

escala de valor da efetividade dos comportamentos observados contra os exemplos providenciados. O critério de seleção baseado na extensão de concordância entre os avaliadores determina a rejeição de um grande número de itens, que são potencialmente comportamentos observáveis. Isto é devido ao procedimento de elaboração das BARS (DeCotiis, 1978). Uma alternativa para estas escalas tem sido desenvolvida para resolver este problema. É o chamado *Behavior Description Index*, onde Schwab (1978) utilizou a Técnica do Incidente Crítico como em BARS mas mudou os incidentes do mais inefetivo para extremamente efetivo.

O critério utilizado na reformulação dos incidentes na elaboração das BARS é que os incidentes são retidos apenas quando a concordância entre os reformuladores atinge uma certa porcentagem, a qual é de 80% para alguns autores e somente 60% para outros, o que leva à confusão, uma vez que nem toda ambiguidade é descartada dos incidentes retidos (Schwab et al., 1975). Os mesmos autores também apontaram para o problema criado pela compactação *shrinkage* dos incidentes na elaboração das BARS. Como foi notado acima, incidentes descartados levam a um tipo de problema, bem como, isto é necessário para evitar ambiguidades.

A falta de um critério externo de desempenho é um outro problema encontrado nos estudos comparando BARS com métodos alternativos (DeCotiis, 1977). O critério externo deve ser desenvolvido independentemente das medidas das BARS e deve conter parâmetros do desempenho em questão.

Um sistema de avaliação de desempenho deve ser prático e valioso *worthwhile* em termos do tempo e custo das fases de elaboração e aplicação da técnica. Os dados coletados de um sistema de avaliação de desempenho deveriam ser muito úteis para os objetivos da organização. Jacobs et al. (1980) afirmaram que o tempo necessário para a elaboração das BARS é provavelmente maior do que a maioria dos sistemas corretamente em uso. Outros sistemas, tais como as Escalas de Observação Comportamental (BOS) (Latham & Wexley, 1981), as escalas numéricas *Numerically Anchored Rating Scales* ou mesmo as escalas com traços (DeCotiis, 1977) são preferidas pelos avaliadores, que explicam sua preferência em termos de praticabilidade, facilidade de entendimento, e assim por diante. Em uma revisão de estudos, Kingstrom & Bass (1981) encontraram que as BARS não eram vistas como o método mais favorável. O tempo consumido e o custo do processo de elaboração das BARS não parece produzir melhores resultados que sistemas desenvolvidos mais facilmente, tais como, escalas gráficas (Bass & Barrett, 1981). Campbell et al. (1973) e Borman & Dunnette (1975) também afirmaram que as BARS requerem muito tempo e esforço para sua elaboração.

Recentes e extensas revisões da literatura (Jacobs et al., 1980; Kingstrom & Bass 1981; Schwab et al., 1975) não encontraram evidências que confirmassem as vantagens hipotetizadas pelos utilizadores das BARS, o que levaram os autores a questionar a justificativa para o investimento de tempo, esforço e dinheiro gasto na elaboração do sistema das BARS. Bernardin & Smith (1981) tentaram justificar o tempo gasto na elaboração das BARS, explicando que o tempo necessário para a elaboração das BARS é um tempo bem gasto, tendo em vista ►

que para documentar e justificar as escalas numéricas e para comparar BARS com outras escalas que consomem muito tempo, tais como as escalas de pontos somados (ver também Landy Farr, 1980).

Jacobs et al. (1980) também sugeriram que qualquer sistema de avaliação de desempenho deveria atingir o critério de disponibilidade; isto é, a informação deveria ser baseada na observação direta do desempenho, embora, mesmo assim, possa haver inferências. Dados adicionais relacionados com o desempenho deveriam também ser disponíveis. Os mesmos autores concordam que as BARS são baseadas na observação e inferência mas que “não há maneira de saber se as inferências são baseadas em observações, situações planejadas ou halo”, uma vez que a literatura tipicamente não inclui informações sobre o uso de documentações.

Bass & Barrett (1981) apontaram que as BARS avaliam executores em comportamentos esperados através de inferências relacionadas ao desempenho dos executores que talvez nunca tenham sido observadas. Por outro lado, Latham & Wexley (1977) notaram a possibilidade do comportamento do executor não parecer com nenhuma dimensão de desempenho da escala. O avaliador deve extrapolar dos comportamentos observados para os esperados.

Afirma-se que uma das maiores vantagens das BARS é o alto nível de participação das pessoas que utilizam o instrumento. A suposição é que, com um alto nível de participação, a terminologia tem mais significado, não é ambígua e motiva os utilizadores a serem mais efetivos e honestos avaliadores (Latham & Wexley, 1981; Schwab et al., 1975). Bass & Barrett (1981), contudo, não concordam com esta suposição e afirmam que, depois de muitos anos, os utilizadores frequentemente não são os mesmos que participaram na elaboração das escalas, uma vez que as pessoas são promovidas, transferidas ou simplesmente deixam o emprego. Keaveny & McGann (1975) assumiram que a participação na elaboração das escalas não tem um grande impacto nos resultados (ver também Bernardin et al., 1976). Depois de rever muitos estudos, Kingstrom & Bass (1981) concluíram que os resultados sobre os efeitos benéficos da participação dos avaliadores na elaboração das BARS não são conclusivos.

Jacobs et al. (1980) concluíram que “não há conhecimento se as BARS são mais, menos ou mesmo adequadas como uma estratégia de avaliação de desempenho quando comparadas com outros sistemas” e que “BARS não são melhores ou piores que outros métodos alternativos de avaliação”. Schwab et al. (1975) concluíram que os resultados de estudos de revisão de literatura “desencorajam futuras pesquisas sobre as BARS”. Além disso, as vantagens deste método nem sempre foram sustentadas pela evidência científica (Bass & Barrett, 1981).

Haynes (1978) criticou a Técnica do Incidente Crítico por ser onerosa e impopular, uma vez que é baseada em comportamentos extremos em vez de desempenho do dia-a-dia, o que revela a medida real da efetividade na situação de trabalho. Futuras pesquisas deveriam estudar a identificação de dimensões do desempenho baseadas na descrição da análise de tarefa (Allan & Rosenberg, 1978; DeCotiis, 1977).

Alguns dos problemas relacionados com as BARS

podem ser atribuídos também às escalas chamadas BOS. Há poucos estudos, entretanto que avaliam a efetividade das BOS. Latham et al. (1979) sugeriram que o uso das BOS para avaliação de desempenho tem muitas vantagens: sua elaboração usa dados fornecidos pelos próprios utilizadores; tem validade de conteúdo; podem ser aplicadas para descrições de tarefa e *feedback* para o desempenho; satisfaz as exigências da EEDC *Equal Employment Opportunity Commission* em termos de confiabilidade e validade; e, finalmente, as BOS evitam alguns problemas encontrados na aplicação das BARS.

Um problema dos sistemas de avaliação de desempenho baseados nas BARS ou BOS é que comportamento e desempenho são vistos como elementos idênticos na análise de tarefa. O comportamento é uma parte importante do desempenho, mas o que a pessoa realiza é também uma medida do desempenho (Gilbert, 1978). Um sistema de avaliação de desempenho deve discriminar o mais eficiente do menos eficiente entre os executores, isto é, deve ter validade discriminante. Medindo comportamentos, um sistema pode não atingir este tipo de validade; segundo Gilbert (1978), “as pessoas são muito parecidas nos repertórios comportamentais... mas existe uma grande diferença naquilo que as pessoas realizam ou produzem”.

Kondrasuk (1981) examinou 185 estudos sobre a eficácia do MBO, estabelecimento de objetivos e gerência pelos resultados na produtividade e/ou satisfação dos executores. O autor mostrou que 153 estudos tiveram resultados positivos, 15 confusos e 17 negativos. Entretanto, somente 33 destes estudos apresentaram dados objetivos coletados na produtividade e/ou satisfação. Dos 185 estudos, aproximadamente 141 eram estudos de caso, 12 eram levantamento de dados, 27 eram quase-experimentos e somente 5 foram considerados experimentos controlados. Entre estes estudos, o mesmo autor também descobriu que os pesquisadores não concordam com a definição de MBO ou não dão informação suficiente sobre esta definição.

SUMÁRIO

O presente estudo mostrou que as técnicas atualmente utilizadas para a avaliação de desempenho são questionáveis para os usos propostos. Problemas relacionados com precisão, confiabilidade, facilidade de execução e validade de conteúdo e discriminante foram encontrados nestes sistemas. Alguns estudos sugerem a necessidade do desenvolvimento de novas abordagens para avaliar o desempenho. De acordo com Kane & Lawler (1979), os sistemas de avaliação existentes são inadequados para a maioria senão para todos os objetivos que a avaliação se propõe a servir, o qual enfatiza a necessidade de desenvolver outros sistemas. Bernardin & Kane (1980) também sugeriram que as pesquisas deveriam concentrar em novas abordagens metodológicas para a medida do comportamento humano.

CONCLUSÃO

Parece que nenhuma técnica para avaliação de de- ▶

sempenho é válida para todas as aplicações. Existem algumas técnicas que contém dados válidos para atingir certos critérios e outras técnicas úteis para atingir outros objetivos. A maioria dos objetivos dos sistemas de avaliação de desempenho requer dados precisos e confiáveis sobre o desempenho atual em termos de realizações esperadas e atingidas e comportamentos necessários para atingir os objetivos de uma dada tarefa. Confiabilidade e precisão são mais fáceis de serem atingidas através de eventos observáveis e mensuráveis quando pequena ou nenhuma inferência é permitida.

De acordo com Gilbert (1978), o comportamento é mais difícil de ser medido satisfatoriamente, porque muitas vezes é encoberto e não é facilmente observado. Além disso, é difícil especificar exatamente quais são os comportamentos necessários para um bom executor, uma vez que eles podem se comportar de maneira muito diferente para atingir o produto final de uma dada tarefa. Portanto, um sistema ideal para avaliar o desempenho deveria incluir todas as realizações esperadas como o elemento básico de uma dada tarefa, ao invés de apenas comportamentos.

Há algumas razões porque o desempenho deveria ser descrito em termos de realizações esperadas ao invés de comportamentos: as realizações esperadas são relacionadas diretamente aos objetivos da organização; podem ser transformados em valor para a organização; são mais facilmente aceitas pelos gerentes e elas descrevem adequadamente uma grande variedade de tarefas parecidas mas diferentes. Em adição, os resultados relevantes que um executor deve atingir em uma dada tarefa são os dados mais adequados para certos objetivos, tais como facilitar tomada de decisões relacionadas a salário, contratação, demissão, transferência ou promoção, para nomear algumas, porque estes dados permitem aos gerentes comparar o desempenho de vários executores, baseados nas realizações esperadas e atingidas por cada executor.

As realizações esperadas ou a descrição dos resultados esperados de uma dada tarefa são medidos e avaliados em termos de qualidade, produtividade e custo exigido para um desempenho de valor. O nível esperado para tal desempenho é chamado o desempenho exemplar ou padrão, contra o qual o desempenho atual é comparado. O executor exemplar é o melhor trabalhador, que atingiu o mais alto nível de desempenho. As-

sim, os padrões não são estabelecidos arbitrariamente; por exemplo, os padrões não são tão altos para serem atingidos e não são baseados no desempenho médio, prática esta que reforça a mediocridade (Gilbert, 1978).

Um modelo funcional que descreva todas as realizações esperadas, os requisitos e padrões para uma ou várias tarefas, parece ser uma base ideal para se desenvolver um sistema para avaliação de desempenho relacionada com decisões administrativas. Uma vez que as deficiências específicas de desempenho tenham sido identificadas, medidas de comportamento devem ser obtidas para facilitar o aperfeiçoamento do comportamento do executor, e, em última análise, do desempenho.

Um SAD baseado neste modelo funcional teria como suporte os princípios teóricos apresentados por Gilbert (1978). O autor sugere a análise e avaliação do sistema como um todo, o qual provê o contexto onde o desempenho das pessoas acontece. Em seu livro *Human Competence — Engineering Worthy Performance*, Gilbert apresenta uma abordagem onde a avaliação de desempenho individual é apenas um dos 18 passos para a análise de desempenho da organização. Chamada de “auditoria de desempenho”, a análise consiste na identificação das realizações esperadas, de maneira mensurável, para a organização como um todo, bem como para cada membro da mesma. Isto constitui o primeiro estágio. Como segundo estágio, o desempenho de toda a organização ou da unidade a que o indivíduo faz parte é avaliado. No terceiro e último estágio, as causas da incompetência ou do desempenho inadequado são analisadas, bem como são sugeridos métodos e programas de aperfeiçoamento.

A metodologia adotada por Gilbert é caracterizada pela simplicidade do sistema de análise, utilidade em relação à grande abrangência dos problemas que atinge e coerência na unidade de análise de todos os níveis da organização.

A avaliação de desempenho nesta abordagem se torna mais real à medida que o ambiente onde o indivíduo trabalha também é analisado; o desempenho não é avaliado isoladamente, mas dentro do contexto organizacional. O objetivo deste SAD é mais relacionado ao desenvolvimento e melhor aproveitamento do potencial humano ao invés de objetivar a promoção ou outras decisões administrativas.

Abstract

A literature review of performance appraisal systems showed that the current techniques are questionable for their intended uses. Problems related to accuracy, feasibility, reliability, and content and discriminant validity were found in those systems. It also appeared that no technique for performance appraisal was suitable for all applications of performance appraisal; some yielded valuable data for meeting certain criteria, and others were useful for meeting other objectives. The literature included suggestions for the development of new approaches for performance appraisal. ▶

Uniterms:

- performance appraisal systems.
- critical review of performance appraisal techniques.
- BOS — behavioral observations scales.
- BES — behavioral expected scales.
- MBO — management by objectives.

Referências Bibliográficas

- ALLAN, Peter & ROSEMBERG, Stephen. The development of a task-oriented approach to performance evaluation in the city of New York. *Public Personnel Management*, Washington, 7(1) 26-32, Jan/Feb. 1978.
- AMERICAN PSYCHOLOGICAL ASSOCIATION. *Principles for the validation and use of personnel selection procedures*. 2 ed. Berkeley: Author, 1980.
- ATKIN, Robert S. & CONLON, Edward J. Behaviorally anchored rating scales: some theoretical issues. *Academy of Management Review*, Mississippi, 3(1):119-28, January 1978.
- BASS, B.M. & BARRETT, G. V. *People, work and organization: An introduction to industrial and organizational psychology*. Boston: Allyn and Bacon, 1981.
- BEATTY, Richard W. et alii. An empirical investigation of perception of ratee behavior frequency and ratee behavior change using behavioral expectation scales (BES). *Personnel Psychology*, Durham, 30(4): 647-58, Winter 1977.
- BEER, Michael et alii. A performance management system: research, design, introduction and evaluation. *Personnel Psychology*, Durham, 31(3): 505-35, Autumn 1978.
- BERNARDIN, H.J. Behavioral expectation scales versus summated rating scales: a fairer comparison. *Journal of Applied Psychology*, Washington, 62:422-7, 1977.
- BERNARDIN, H. J. et alii. A recomparison of behavioral expectation scales to summated scales. *Journal of Applied Psychology*, Washington, 61: 564-70, 1976.
- BERNARDIN, H. John & KANE, Jeffrey. A second look at behavioral observation scales. *Personnel Psychology*, Durham, 33(4): 809-14, Winter 1980.
- BERNARDIN, H.J. & SMITH, P.C. A clarification of some issues regarding the development and use of behaviorally anchored rating scales (BARS). *Journal of Applied Psychology*, Washington, 66: 458-63, 1981.
- BLOOD, Milton R. Spin-offs from behavioral expectation scale procedures. *Journal of Applied Psychology*, Washington, 59(4): 513-5, August 1974.
- BORMAN, W. C. Exploring upper limits of reliability and validity in job performance rating. *Journal of Applied Psychology*, Washington, 63: 135-44, 1978.
- BORMAN, W. C. Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, Washington 64: 410-21, 1979.
- BORMAN, Walter C. & DUNNETTE, Marvin D. Behavior-based versus trait oriented performance ratings: an empirical study. *Journal of Applied Psychology*, Washington, 60(5): 561-5, October 1975.
- CAMPBELL, John P. et alii. The development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology*, Washington, 57(1): 15-22, February 1973.
- CUMMINGS, L.L. & SCHWAB, D. *Performance in organizations: Determinants and appraisals*. Glenview, Ill., Scott Foresman, 1973.
- DeCOTIIS, T. A. An analysis of the external validity and applied relevance of three rating formats. *Organizational Behavior and Human Decision Processes*, San Diego, CA, 1977, 19: 247-66.
- DeCOTIIS, T. A. A critique and suggested revision of behaviorally anchored rating scales developmental procedures. *Educational and Psychological Measurement*, Durham, 38(3): 681-90, 1978.
- DeCOTIIS, Thomas & PETIT, Andre. The performance appraisal process: a model and some testable propositions. *Academy of Management Review*, Mississippi, 3(3): 635-46, July 1978.
- FLANAGAN, John C. The critical incident technique. *Psychological Bulletin*, Washington 51(4): 327-58, July 1954.
- GILBERT, T. F. Praexeonomy: a systematic approach to identifying training needs. *Management of Personnel Quarterly*, 6(3): 20-33, 1967.
- GILBERT, T. F. *Human competence*, New York, McGraw-Hill, 1978.

- HAYNES, Marion G. Developing an appraisal program. *Personnel Journal*, Califórnia, 1978, **57**(1): 14-9, January 1978.
- HANDERSON, R. I. *Compensation management: Rewarding performance in the modern organization*. Reston, Va, Prentice-Hall, 1976.
- HENRICI, S. B. *Salary management for the nonspecialists*. New York, AMACOM, 1980.
- IVANCEVICH, J. M. A longitudinal study of behavioral expectation scales: attitudes and performance. *Journal of Applied Psychology*, Washington, **65**: 139-46, 1980.
- JACOBS, Rick et alii. Expectations of behaviorally anchored rating scales. *Personnel Psychology*, Durham, **33**(3): 595-640, Autumn 1980.
- KAHALAS, H. The environmental context of performance evaluation and its effect on current practices. *Human Resource Management*, Ann Arbor, Mich., **19**(3): 32-40, 1980.
- KANE, J.S. & LAWLER, E. E. Performance appraisal effectiveness: its assessments and determinants. In: STAW B. ed. *Research in organizational behavior*. Greenwich, JAI Press, 1979. v.1
- KEAVENY, Timothy J. & McGANN, Anthony F. A comparison of behavioral expectation scales and graphic rating scales. *Journal of Applied Psychology*, Washington, **60**(6): 695-703, December 1975.
- KINGSTROM, Paul O. & BASS, Alan R. A critical analysis of studies comparing behaviorally anchored rating scales (BARS) and other rating formats. *Personnel Psychology*, Durham, **34**(2): 263-89, Summer 1981.
- KONDRASUK, Jack N. Studies in MBO effectiveness. *Academy of Management Review*, Mississippi, **6**(3): 419-30, July 1981.
- KRAUSZ, M. Performance appraisal in a highly unionized country. *Human Resource Management*, **19**(2): 14-8, 1980.
- LACHO, Kenneth J. et alii. A study of employee appraisal systems of major cities in the United States. *Public Personnel Management*, Washington, **8**(2): 111-25, March/April 1979.
- LANDY, F. J. & FARR, J. L. Performance rating. *Psychological Bulletin*, **87**: 72-107, 1980.
- LATHAM, Gary P. et alii. The development of behavioral observation scales for appraising the performance of foremen. *Personnel Psychology*, Durham, **32**(2): 299-311, Summer 1979.
- LATHAM, Gary P. & Wexley, Kenneth N. Behavioral observation scales for performance purposes. *Personnel Psychology*, Durham, **30**(2): 255-68, Summer 1977.
- LATHAM, G.P. & WEXLEY, K. N. *Increasing productivity through performance appraisal*. Reading, MASS., Addison-Wesley, 1981.
- LAWER, E. E. *Pay and organization development*. Reading, MASS., Addison-Wesley, 1981.
- LAZER, R.T. & WIKSTROM, W. S. *Appraising managerial performance: Current practices and futures directions*. New York, S. ed., 1977.
- LEFTON, R. E. & BUZZOTTA, V. R. Performance appraisal: why bother? *Training and Development Journal*, **32**:49-58, 1978.
- LEVINSON, Harry. Appraisal of WHAT performance? *Harvard Business Review*, Boston, **54**(4): 30-46, July/August 1976.
- LOCHER, Alan H. & TELL, Kenneth S. Performance appraisal: a survey of current practices. *Personnel Journal*, California, **56**(5): 245-7, 54, May 1977.
- MCCORMICK, E.J. & TIFFIN, J. *Industrial psychology*. Englewood Cliffs, Prentice-Hall, 1965.
- MacGREGOR, Douglas. An uneasy look at performance appraisal. *Harvard Business Review*, Boston, **35**(3): 89-94, May/June 1957.
- MAYER, H.H. et alii. Split roles in performance appraisal. *Harvard Business Review*, **43**: 123-29, 1965.
- NADLER, D.A. et alii. The ongoing feedback system: experimenting with a new managerial tool. *Organizational Dynamics*, **4**: 63-80, 1976.
- OBERG, WINSTON. Make performance appraisal relevant. *Harvard Business Review*, Boston, **50**(1): 61-7, January/February 1972.
- ODIORNE, G. S. *M.B.O. II*. Belmont, Calif., Fearon, 1979.
- RALPH, P. M. Performance evaluation: One more try. *Public Personnel Management*, Washington, **9**(3): 145-53, 1980.
- RUMMLER, G. A. Human performance problems and their solutions. *Human Resource Management*, Ann Arbor, Mich, **11**(4):2-10, 1972.
- SANDERS, M.S. & PEAY, J. M. *Employee performance evaluation and review: a summary of the literature*. IN: Crane, Naval Ammunition Depot, 1974.
- SCHWAB, Donald P. et alii. Behaviorally anchored rating scales: a review of the literature. *Personnel Psychology*, Durham, **28**(4):549-562, Winter, 1975.
- SCHWIND, H. F. An alternative to behaviorally anchored rating scales: the behavior description index. *Academy of Management Proceedings*, 1978, **5**: 38-42.
- SIKULA, A. F. & SIKULA, J. P. Rethinking present appraisal systems. *Supervisory Management*, **23**: 12-6, 1978.
- SMITH, Patricia Cain & KENDALL, L. M. Retranslation of expectations: an approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, Washington, **47**(2):149-155, 1963.
- SMITH, H.P. & BROWVER, P.J. *Performance appraisal and human development*. Reading, Mass., Addison-Wesley, 1977.

- VROOM, V. W. & MAIER, N. R. F. Industrial social psychology. *Annual Review of Psychology*, 12:413-46, 1961.
- WARREN, M. W. *Training for results* 2 ed. Reading, Mass, Addison-Wesley, 1979.
- WEXLEY, K.N. & YUKL, G. A. *Organizational behavior and personnel psychology*. Homewood, Ill., Irwin, 1977.
- WILLIAMS, M. R. *Performance appraisal in management*. London, Heinemann, 1972.
- YAGER, Ed. A critique of performance appraisal systems. *Personnel Journal*, California. 60(2):129-33, February 1981.
- ZEDECK, S. et alii. Format and scoring variations in behavioral expectations evaluations. *Organizational Behavior and Human Decision Processes*, San Diego, CA, 17: 171-84, 1976.

Recebido em novembro/87

2ª versão em maio/88