

**Os testes psicológicos em seleção de pessoal:
análise crítica dos conceitos e procedimentos utilizados***

WEDHER MODENEZI WANDERLEY**

“Meu objetivo é ajudá-los a passar de um absurdo disfarçado para algo que seja patentemente absurdo” (Wittgenstein)

1. As expectativas do indivíduo submetido ao processo de seleção; 2. Resultados obtidos com os testes; 3. A utilização das aptidões e seu significado; 4. A necessidade de comprovação das hipóteses de seleção; 5. Redefinição do conceito de validade dos testes; 6. Os quatro tipos de validade; 7. Conclusão.

Este artigo questiona conceitos e procedimentos que orientam o emprego dos testes psicológicos em seleção de pessoal. Examina levantamentos de pesquisas realizadas com testes para as ocupações, concluindo pela inexpressividade dos resultados. Considera irreal a interpretação dada às características psicológicas e insiste na comprovação do relacionamento teste-desempenho antes de o preditor ser utilizado para fins decisórios. Analisa a inadequação dos vários conceitos de validade. Cita restrições ao uso dos testes, impostas, nos EUA, pelo governo e pelas cortes de justiça, mostrando a necessidade de os psicólogos brasileiros elaborarem normas rigorosas para o emprego de testes, antecipando-se às restrições legais que inevitavelmente surgirão.

1. As expectativas do indivíduo submetido ao processo de seleção

Ao examinarmos o processo de seleção de pessoal, precisamos considerar que, no momento de candidatar-se a um cargo em qualquer organização, o indivíduo confia em que serão julgadas, na admissão, suas reais possibilidades de realizar as tarefas ali previstas. Espera ser avaliado a partir de critérios cla-

* Artigo apresentado à Redação em 22.10.84.

** Psicólogo do ISOP, coordenador do Curso de Especialização (pós-graduação) em Psicologia do Trabalho. (Endereço do autor: Rua Constante Ramos, 78/303 — Copacabana — 22.051 — Rio de Janeiro, RJ.)

ramente explicitados e, se está suficientemente preparado, saberá que solicitações lhe serão feitas, devendo esse "saber antecipado" ser tanto maior quanto melhores forem suas condições de exercer a função. Entretanto, quando a esse indivíduo se nega a possibilidade de ocupar o cargo, em virtude de razões que escapam ao seu entendimento e que não estão comprovadamente relacionadas, de forma decisiva, com o desempenho no trabalho, estão-se comprometendo os seus interesses, bem como os interesses da organização e até os da própria sociedade, sob aspectos econômicos, sociais, morais e humanos.

Para ser compatível com tais pressupostos, que orientam a tomada de decisões referentes à admissão ou rejeição de candidatos às organizações, a utilização dos testes psicológicos deve estar firmemente amparada em estudos que não deixem dúvidas quanto à estreita relação entre o que é medido pelo teste e o que é realizado no trabalho. E isso vale em qualquer situação, seja para datilógrafos candidatos a cargos administrativos, seja para psicólogos candidatos ao setor de recursos humanos de uma empresa. Entretanto, os estudos específicos, visando demonstrar esses relacionamentos, raramente têm conseguido atingir seus objetivos. Existem mesmo vários trabalhos abrangentes, mostrando que as expectativas históricas em relação aos resultados da utilização dos testes psicológicos em seleção não se concretizaram.

2. Resultados obtidos com os testes

Em artigo publicado em 1973, Ghiselli (1973) fez o resumo dos resultados obtidos com testes ocupacionais nos EUA, entre 1920 e 1971, representando todas as pesquisas dessa área, publicadas nesse período, além de grande quantidade de material não-publicado oriundo de trabalhos realizados em organizações comerciais, industriais e governamentais. Categorizou os testes estudados em 20 classes, os cargos em 21 categorias e tomou, como critérios, o nível de eficiência no cargo e/ou o sucesso no treinamento. Expressando a validade em termos de correlação entre o teste e o critério, ele calculou as médias dos coeficientes de validade para cada uma das classes de teste e cada uma das categorias de cargos, em relação aos dois tipos de critério, separadamente. A média global desses coeficientes para todas as modalidades de testes e cargos tomados em conjunto foi de 0,22, em relação aos critérios de eficiência nos cargos, o que importa numa variância comum teste-critério inferior a 5%, índice que, a nosso ver, dificilmente pode ser considerado suficiente, particularmente, para decisões individuais. Entre a grande quantidade de material relatado por Ghiselli, cumpre observar um dado que nos parece muito significativo, como noção de limite: separando o mais alto coeficiente de validade média para cada uma das 21 categorias de cargos, seus valores, para os critérios de eficiência, variaram entre os limites de 0,24 e 0,46, ficando a média desses coeficientes máximos em 0,35.

As validades médias, obtidas por Ghiselli, para todos os testes classificados em cinco grandes classes, em relação aos cargos categorizados em oito grandes grupos, no critério de eficiência (quadro 1), demonstraram, em nossa opinião, que não dispomos de bases razoáveis, para afirmações genéricas sobre as potencialidades preditoras das várias classes de testes em relação aos vários grupos de cargos. Observe-se que a maioria dessas validades médias tem muito boa estabilidade, conforme se pode concluir pelo número de casos utilizados em cada uma.

Quadro 1

Coefficientes de validade média obtidos entre classes gerais de testes e desempenho em grandes categorias ocupacionais

Tipos de testes	Ocupações							
	Gerentes em geral	Funções burocráticas em geral	Ocupações de vendas em geral	Ocupações de segurança em geral	Ocupações de serviços em geral	Operadores de veículos em geral	Comerciantes e artesãos em geral	Ocupações industriais em geral
Habilidades intelectuais	0,27 ^f	0,28 ^f	0,19 ^e	0,22 ^d	0,27 ^d	0,16 ^f	0,25 ^f	0,20 ^f
Habilidades espaciais e mecânicas	0,22 ^e	0,17 ^f	0,18 ^e	0,18 ^d	0,13 ^d	0,20 ^d	0,23 ^f	0,20 ^f
Acuidade perceptiva	0,25 ^c	0,29 ^f	0,04 ^d	0,21 ^c	0,10 ^d	0,17 ^b	0,24 ^e	0,20 ^f
Habilidades motoras	0,14 ^d	0,16 ^f	0,12 ^d	0,14 ^d	0,15 ^d	0,25 ^d	0,19 ^f	0,22 ^f
Características de personalidade	0,22 ^f	0,22 ^d	0,32 ^e	0,21 ^c	0,16 ^b	0,26 ^b	0,24 ^c	0,26 ^a

Fonte: Dados extraídos do estudo de Ghiselli (1973).

^a Menos de 100 casos.

^b De 100 a 499 casos

^c De 500 a 999 casos.

^d De 1 mil a 4.999 casos.

^e De 5 mil a 9.999 casos.

^f Dez mil ou mais casos.

Embora, para os objetivos de nossa análise, os resultados referentes aos critérios de eficiência no cargo sejam os mais importantes, devemos ressaltar que as validades médias para os critérios de treinamento foram, de modo geral, bem melhores. Assim, a validade média global obtida para esse tipo de critério foi de 0,39, o que se explica em face da existência de semelhanças entre o conteúdo dos testes e os conhecimentos, de tipo escolar, mobilizados pelos cursos de treinamento.

Em 1972, Dunnette (1976) e outros apresentaram resumo de todos os trabalhos de validação até então publicados, bem como de muitos não-publicados, referentes às funções relevantes da indústria de refinação de petróleo (com exceção de supervisores). Embora Dunnette considere os coeficientes de validade média obtidos superiores aos registrados por Ghiselli, eles, geralmente, continuam muito abaixo do aceitável para uso na contratação de pessoal, como se pode observar no quadro 2, onde os coeficientes relatados para as funções burocráticas e para as funções de controle de qualidade não ultrapassaram 0,24, em todas as classes de testes de aptidão utilizados.

Quadro 2

Coefficientes de validade média em várias áreas de aptidão relativas ao desempenho em quatro setores ocupacionais relevantes para a refinação do petróleo

Area de aptidão	Setor ocupacional			
	Operação e processamento	Manutenção	Burocrática	Controle de qualidade
Inteligência geral	0,32(81)	0,20(11)	0,17(14)	0,24(8)
Numérica	0,19(36)	0,35(86)	0,12(31)	0,14(10)
Verbal	—	0,29(16)	0,22(8)	0,16(4)
Habilidade mecânica	0,20(42)	0,38(76)	—	0,18(8)
Velocidade perceptiva	—	0,16(43)	0,15(5)	—
Relações espaciais	—	0,24(34)	0,04(10)	—
Habilidades motoras	0,17(9)	0,22(65)	0,22(19)	0,06(2)

Fonte: Dunette (1976, cap. 11, p. 499).

Obs.: os números entre parênteses representam o número total de coeficientes utilizados para o cálculo.

Guion & Gottier (1965) fizeram longo levantamento dos resultados das pesquisas sobre o emprego das medidas de personalidade em seleção de pessoal, para empregos civis, publicados no *Journal of Applied Psychology* e no *Personnel Psychology*, de 1952 a 1963, chegando à desalentadora constatação de que “não se pode dizer que qualquer das medidas de personalidade convencionais tenha demonstrado utilidade real como instrumento de seleção” e concluem: “algumas vezes são empregadas medidas de personalidade não-validadas como instrumento de decisão, devido a ‘insights clínicos’, credulidade, superstição ou evidência acumulada em algum cenário. Todas elas podem ser igualmente condenadas, a menos que se consiga reunir dados específicos da

situação, revelando que os *insights*, superstições ou validades são de fato preditivas." Lembrem, ainda, o levantamento das validades dos testes de personalidade, no meio industrial, feito por Ghiselli e Barthol, em 1953, em que esses autores constataram correlações mais fracas justamente nos grupos ocupacionais em que os traços de personalidade pareciam intuitivamente mais importantes.

De modo geral, os mais considerados estudos feitos com testes de personalidade criticam de forma eloqüente, no mínimo, as fracas qualidades psicométricas desses testes, particularmente daqueles classificados como projetivos ou expressivos, contra-indicando-os para uso em seleção de pessoal; entretanto, em relação aos testes de aptidão não existe convergência de pontos de vista entre os principais psicólogos que realizaram trabalhos nessa área, como Guion, Faverge, Ghiselli, Montmollin, Cronbach, Ebel, Lawshe e outros. A verdade é que a fragilidade dos conceitos e procedimentos aí utilizados, até por questões de envolvimento pessoal, tem sido constantemente ignorada. Mas há os resultados; e eles, conforme indicamos, são inexpressivos, após tantos anos de uso dos testes psicológicos. Isto é um fato. Clark Hull já havia observado, em 1928, que o teto de validade para os testes ficaria em torno de 0,50. De lá para cá não houve melhoras.

Se muitos psicólogos consideram desconcertante que esses resultados venham ocorrendo e acham que a única solução é trabalharmos para aperfeiçoar os modelos e os testes empregados, alguns poucos, como nós, julgam até estranho que se prossiga atuando na mesma direção, como se ainda estivéssemos explorando filão de grande futuro. Chegamos a pensar que é um grande despropósito continuarmos a desenvolver tão sofisticada maquinaria estatística, abastecida pela excelente tecnologia a esse respeito disponível, para produzir tanto ruído e tão pouco conhecimento significativo no sentido dos objetivos desejados.

Então, ponderamos: não estamos alcançando resultados úteis; não estamos produzindo conhecimento científico aceitável; não estamos conseguindo compatibilizar a decisão de usar testes psicológicos com os objetivos mais sadios das políticas de pessoal das organizações. O que estamos fazendo?

Na minha opinião, esse desencontro entre aquilo que pretendemos fazer e o que realmente estamos fazendo em seleção de pessoal é consequência de não termos questionado, suficientemente, ao longo desses anos, alguns conceitos e procedimentos aparentemente lógicos que dominam o conhecimento nessa área.

3. A utilização das aptidões e seu significado

Assim, por exemplo, muitos psicólogos costumam deduzir rapidamente, até por simples leitura dos deveres funcionais, características ditas psicológicas necessárias ao desempenho no cargo, como: aptidão espacial, para mecânicos; aptidão verbal, para jornalistas; extroversão, para vendedores; coordenação bimanual, para motoristas; estabilidade emocional para chefia e assim por diante. E o que há de errado nisso? Bem, em primeiro lugar, o sucesso no trabalho não pode ser explicado apenas por fatores pessoais de aptidão, personalidade etc., e sim, pela conjunção de competência e utilização de certas condições pessoais, com uma situação particular. Não se pode afirmar que uma pessoa

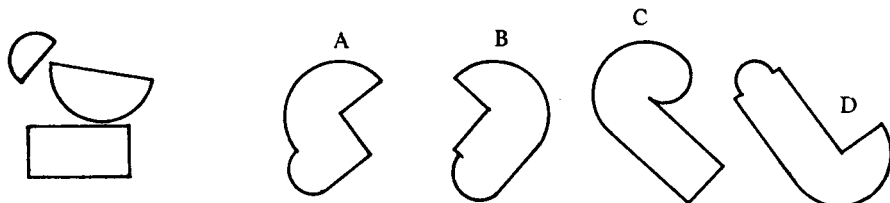
para ter êxito em posição de chefia deva possuir, por exemplo, muito boa estabilidade emocional; a história está aí para demonstrar a irrealidade dessa posição. Além disso, o sucesso no desempenho do trabalho, dentro de qualquer organização, vai depender, também, das circunstâncias organizacionais em que as tarefas forem realizadas, como: tipo de chefia, incentivos comportamentais existentes, condições de trabalho, clima organizacional etc. Mesmo em igualdade de condições, pessoas possuidoras de características básicas bem diferentes, até opostas, podem ter o mesmo êxito utilizando-as adequadamente, dentro das circunstâncias organizacionais. Em segundo lugar, conforme demonstrou Maurice Montmollin (1974), a linguagem utilizada para descrever características individuais não é a mais adequada para descrever as tarefas do trabalho. Para Montmollin, quando descrevemos os cargos empregando termos como inteligência, fluência verbal, raciocínio geral etc., correspondentes aos nomes de fatores isolados pela psicologia diferencial, estamos usando um truque maroto tornando "evidente" a indicação automática dos testes psicológicos a serem utilizados para esses cargos, pois não é por coincidência que esses testes pretendem medir aqueles fatores. No Brasil essa maneira de analisar o trabalho caracteriza o que se chama análise profissiográfica.

Tudo isso torna ilógico projetarmos, a partir desses dados, critérios comportamentais de desempenho no trabalho. Mas há também outra questão mais profunda: quando consideramos que uma característica psicológica é decisiva para o desempenho em alguma tarefa e que podemos medi-la através de testes, estamos, de forma implícita, acreditando que existem certos traços, constructos, processos mentais, ou seja lá o que for, por trás tanto dos escores do preditor como do desempenho das tarefas do trabalho, governando a atuação do indivíduo, e que os esforços para medir esses constructos não são mais do que aproximações, em relação a alguma coisa que não temos condições de definir operacionalmente, porém se continuarmos a investigar nessa direção e aperfeiçoarmos nossos instrumentos de medida, talvez um dia se chegue lá. Mas será que a psicologia, no estágio atual do seu desenvolvimento, tem condições de explicar constructos, como aptidão espacial, aptidão numérica ou aptidão verbal em termos, por exemplo, de processos mentais utilizados?

O estudo dos processos intelectuais, nesse campo da psicologia, é feito em termos de caixa preta em que as conclusões são obtidas pelo confronto das "entradas" e "saídas" do modelo experimental utilizado. Introduzimos nessas "entradas" certos testes para medir determinadas aptidões porque, pela análise de seus conteúdos, "parece evidente", em termos de senso comum, que eles medem aquilo a que se propõem (Wanderley, 1976). Assim, um teste é de aptidão espacial porque se supõe que, para a resolução de suas questões, uma pessoa deva, por exemplo, manipular mentalmente no espaço de três dimensões representações esquemáticas de determinados objetos. De forma semelhante, se os itens do teste consistirem na escolha de sinônimos e antônimos, ele será considerado de aptidão verbal; se os itens implicarem a realização de contaras das quatro operações, será um teste de aptidão numérica. Então, o que caracteriza realmente a aptidão que o teste mede é o tipo de material utilizado — nos casos citados: figurativo, verbal e numérico. A importância do material empregado nos testes tem sido evidenciada, como lembra Faverge (1972), "pelos trabalhos clássicos de análise fatorial em que, contra toda a previsão e, por vezes, com consternação, emergem na primeira linha, esgotando a maior parte da variância comum, os fatores espaciais, verbais e numéricos".

Examinemos um pouco mais a questão. Consideremos três ocupações diferentes para as quais geralmente se supõe que a aptidão espacial é importante: mecânico de carburador de automóvel, alfaiate e mestre-de-obras de construção civil. Suponhamos que devemos planejar a seleção de candidatos para essas ocupações e que dispomos dos três tipos de testes de aptidão espacial apresentados na figura 1. Se tivermos de escolher, entre esses testes, os que serão incluídos nas baterias de seleção para cada uma das ocupações, como faremos a escolha? Por quê?

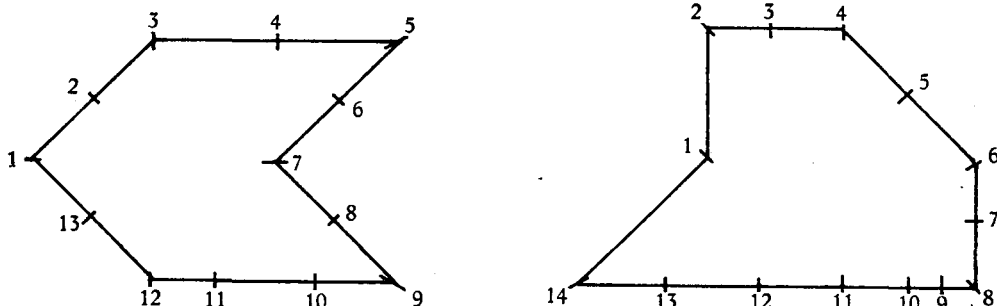
Figura 1



Teste A. Qual das figuras é formada pela junção das partes da questão apresentada à esquerda?



Teste B. Quantos cubos tem cada uma das figuras?



Teste C. Corte a figura em duas partes, através de uma linha reta passando por dois pontos numerados, de modo que essas partes possam formar um quadrado.

Acredito que a tendência do psicólogo, não dispondo de outras informações, será indicar o teste A para os mecânicos, o teste C para os alfaiates e o teste B para os mestres-de-obras. Parece fora de dúvida que essa escolha

seria determinada pela possível semelhança entre as tarefas de cada teste e as tarefas a serem realizadas nessas ocupações, ou seja, pela manutenção, na medida do possível, da relação sujeito-objeto. Embora esses testes tenham entre eles certas semelhanças por utilizarem material figurativo, as maneiras de resolver as respectivas questões são um pouco diferentes e, se quiséssemos fazer suposições sobre as operações mentais envolvidas, não poderíamos deixar de pensar nessas diferenças. Por outro lado, a aptidão espacial por eles representada não pode ser considerada a mesma nos três testes apresentados, pelo menos, não no sentido que estamos tratando. Na hipótese contrária, isto é, se pudéssemos concluir, através desse caminho, pela existência de aptidão bem definida, em termos, talvez, de processo mental responsável pela resolução dos citados testes, bastaria definirmos o teste que melhor representasse esse processo, para utilizá-lo, daí por diante, toda vez que ficasse caracterizada a importância daquela aptidão.

Mas, se essas aptidões não representam certas "faculdades" ou processos intelectuais que a pessoa utiliza para resolver as questões de teste e realizar as tarefas do trabalho, o que significa terem elas aparecido nas análises fatoriais realizadas?

A resposta é que um fator obtido na análise fatorial é apenas uma entidade estatística. A interpretação no sentido de considerá-lo fator psicológico, assim como a identificação do mesmo fator em vários conjuntos diferentes de testes submetidos às análises fatoriais, constituem hipóteses do pesquisador. Em termos objetivos, um fator indica, apenas, o caráter comum dos diferentes testes submetidos a essa técnica e, desse modo, o seu significado não pode ir além do conteúdo dos testes ali incluídos.

Thurstone, um dos primeiros a utilizar a análise fatorial, para identificar dimensões comportamentais, observou, conforme resumo de Bechtoldt (1959), que "uma capacidade é definida por um procedimento específico de teste e pelo método de qualificação (Thurstone). Cada teste define uma capacidade em separado e vários testes, que incluem o mesmo conteúdo, porém, diferentes métodos de apresentação dos estímulos ou das respostas, também definem diferentes capacidades". Esta linha de pensamento é também endossada por Robert Ebel (1961), do Educational Testing Service, ao criticar os psicólogos que acreditam na existência de características humanas quantificáveis, como a inteligência, independente da operação que as meça, ou seja, independente de qualquer definição operacional, como a de Binet ou de Wechsler, por exemplo.

Talvez fosse melhor deixarmos de falar, de uma vez por todas, nesses constructos em seleção de pessoal. Não seria muito mais correto dizermos que um teste verifica o conhecimento de sinônimos e antônimos, ao invés de declararmos que ele mede aptidão verbal? É justo concluirmos que alguém possui baixo nível de aptidão científica, porque não se saiu bem num teste de problemas engenhosos, como o de Zyve?

4. A necessidade de comprovação das hipóteses de seleção

Existe um abuso nesses saltos de interpretação; entretanto, a grande maioria dos psicólogos ainda não pensa dessa maneira. Consideram que a realização nos testes e a execução das tarefas do trabalho refletem a mobilização de

certos atributos pessoais, conforme indicamos anteriormente, e tomam decisões sobre a admissão de candidatos aos cargos a partir dessa idéia. Esquecem os nossos companheiros que essa forma de pensar se traduz, em cada processo seletivo, em hipóteses que precisam ser comprovadas *antes* de serem utilizadas; caso contrário, continuarão sendo apenas crenças, como outras quaisquer.

Robert Guion (1976) mostra que, já em 1923, Freid havia elaborado o modelo de pesquisa em seleção que exigia essa comprovação, de acordo com as normas da metodologia científica ortodoxa, estabelecendo a seguinte seqüência de fases a serem desenvolvidas no processo: a) análise do trabalho; b) determinação do critério; c) determinação da população alvo para a experiência; d) desenvolvimento de uma lista de habilidades necessárias ao sucesso; e) construção ou escolha dos instrumentos adequados (não se restringindo aos testes); f) aplicação sob condições controladas; g) comparação estatística dos resultados do teste com os escores do critério; h) combinação das medidas para melhorar a prognose; i) comparação da predição dos novos instrumentos com a já em uso; j) se favorável, adoção dos novos instrumentos; l) verificação constante de todas as fases.

Havia nesse modelo dois elementos básicos: de um lado, o critério; do outro, o instrumento de prognose, geralmente o teste. Como o critério para avaliar o desempenho era o elemento a ser predito, ele precisaria ser elaborado em primeiro lugar e isso teria que ser feito a partir da análise do trabalho. Tendo o critério, este daria origem à construção ou escolha do instrumento de medida para predizê-lo. Somente depois de comprovada a relação teste-critério e verificada a superioridade de prognose dos novos preditores em relação aos já em uso é que eles poderiam ser utilizados. Os psicólogos tinham, portanto, um bom modelo para as suas pesquisas em seleção de pessoal.

A utilização sistemática dessa metodologia, desde a sua implantação, provavelmente nos teria despertado, há mais tempo, para a irreidade de nossas suposições sobre os testes e os fatores. Entretanto, embora o modelo não fosse rejeitado, foi silenciosamente esquecido; lembra Robert Guion (1976) que lá pelo final dos anos 50 a seleção de pessoal começou a desviar-se do rigoroso modelo que exigia a comprovação de todos os resultados, pois “todos sabiam que os testes eram maravilhosos. (...) Os não-psicólogos não tinham consciência das muitas coisas que podiam sair erradas num programa de testes e os psicólogos estavam freqüentemente muito ocupados com o recém-aumentado escopo de sua disciplina para se preocuparem com isso”.

Alguma coisa precisava acontecer para que fossem corrigidos os rumos que a seleção de pessoal havia tomado. Certa vez Mark Twain disse: “Uma coisa é ler que não se deve carregar um gato pelo rabo. Para aquele que o tentou, no entanto, a lição é tão vívida quanto difícil de esquecer.”

Para os nossos colegas norte-americanos e para as organizações a que serviam, a lição chegou quando, nos últimos anos, pessoas que, de alguma forma se julgaram prejudicadas pelos testes psicológicos, em seu direito ao emprego, recorreram à justiça. Creio que vale a pena tomarmos conhecimento de alguns problemas que os psicólogos norte-americanos passaram a enfrentar nas cortes de justiça, pois a maioria deles, até onde estamos informados, refere-se à rejeição de conceitos e procedimentos inadequados que vinham sendo utilizados. Por outro lado, as mudanças que estão ocorrendo naquele país, em direção a uma tecnologia de seleção mais rigorosa do ponto

de vista científico e mais compatível com as reais possibilidades da psicologia, deverão refletir-se, como não pode deixar de ser, em modificações de nossa própria forma de atuar.

Nos EUA, esses problemas começaram com as previsões constantes do Ato de Direitos Civis de 1964 relativas à eliminação da discriminação racial, étnica, religiosa ou sexual. Nessa ocasião, foi considerado que os testes psicológicos podiam funcionar como barreiras às oportunidades de emprego das minorias. Isso levou uma Comissão de Iguais Oportunidades de Emprego (EEOC) a publicar diretrizes obrigando o empregador a validar qualquer procedimento de seleção que resultasse em escolha desproporcionalmente baixa de minorias ou de mulheres. A partir das revisões desse Ato de Direitos Civis, publicadas em 1972, e das sucessivas regulamentações sobre o uso de testes feitas pela Comissão de Iguais Oportunidades de Emprego (EEOC) e pelo Departamento de Trabalho daquele país, o controle do governo federal passou a realizar-se não apenas em relação à discriminação das minorias, mas também sobre todas as outras questões inerentes ao processo de seleção de pessoal, envolvendo também questões técnicas, até então da exclusiva alçada dos psicólogos.

A intervenção governamental na área técnica foi apoiada por várias decisões judiciais, inclusive da Suprema Corte dos EUA, estabelecendo coisas como: padrões e características de uma análise do trabalho para que tivesse valor legal na validação de testes (Thompson, 1982), parâmetros que deviam ser utilizados para examinar a validade de conteúdo de um teste, conceitos de validade que não deveriam ser considerados etc. Como exemplo do tipo de interferência externa nesses assuntos técnicos, podemos citar a decisão de uma corte distrital de Nova Iorque, em 1979 (Thompson, 1982), rejeitando o trabalho de validação do instrumento de medida para exame do policial montado, realizado pelo estado de Nova Iorque com a assistência técnica de Comissão de Serviço Civil dos EUA, em projeto no qual foram gastos US\$ 1.250 mil. A conclusão da Corte foi: "O método de análise de cargos baseada nos elementos de cargo" (trata-se de um método denominado Coeficiente J) "seguido pela NYSP na elaboração, em 1975, do exame para policiais montados, não se concentra naquilo que os cavalarianos realmente fazem em serviço, mas somente nos traços ou nas características subjacentes que os policiais acreditam caracterizar o desempenho bem-sucedido. (...) Não existe qualquer ligação documentada entre o campo de conteúdo do exame e as tarefas reais, deveres e atividades do policial montado do estado de Nova Iorque."

Sucessivas regulamentações governamentais bem como novas decisões das cortes de justiça vêm ampliando cada vez mais o campo de restrições ao emprego dos testes. Assim, atualmente algumas decisões parecem estar muito bem estabelecidas, como, por exemplo, o dever de o empregador demonstrar que um teste usado na contratação é preditor do desempenho no cargo, não sendo aceita a idéia de que ele meça alguma característica psicológica que, apenas supostamente, seja relevante para o cargo. Além disso, e aqui está uma decisão técnica difícil de cumprir, esse relacionamento teste-desempenho deve ser demonstrado para cada cargo de uma companhia, o que implica validação local. Dentro dessa restrição, as empresas "foram proibidas de usar testes cujos estudos de validação tivessem sido executados em outro lugar, mesmo que esses estudos fossem extensa e tecnicamente bem fundamentados" (Resnick, 1982).

A obrigatoriedade de comprovação da relação teste-desempenho, da validação local, além da utilização de medidas comprovadamente não discriminatórias, tornaram, nos EUA, o emprego dos testes psicológicos em seleção algo tão dispendioso que somente as grandes empresas podem realizar esses estudos. Além disso, os resultados ainda estão sujeitos à contestação nas cortes de justiça, o que envolve não só custos elevados de defesa judicial, como alguns prejuízos referentes ao conceito da organização.

5. Redefinição do conceito de validade dos testes

Nesse conflito de posições envolvendo empresas, psicólogos-assessores, queixosos, psicólogos-peritos, advogados etc., regulado pelo poder judiciário através das decisões das cortes de justiça, surgem a todo momento referências à validade ou ao processo de validação dos testes, refletindo, provavelmente, a respeitabilidade quase mágica que esses conceitos adquiriram dentro da psicologia. Entretanto, nós não podemos ter muita certeza sobre o que estão falando diferentes pessoas quando dizem que um teste é válido. É até possível que o apego a esse conceito-força, de significado ambíguo, tenha contribuído, junto com os outros motivos citados anteriormente, para nos enredar num cipoal de confusões a ponto de tornar penoso o progresso do conhecimento nessa área da psicologia do trabalho. Julgo que, se quisermos colocar a seleção de pessoal na direção correta, precisamos também desmontar algumas armadilhas conceituais, como a referente à validade dos testes.

Assim, geralmente entendemos como validade de um teste a sua capacidade de medir aquilo a que se propõe; porém como essa definição não é operacional, passamos a caracterizar a validade pelo coeficiente de correlação teste-critério. Acontece que a validade assim operacionalizada assinala a acuidade das inferências obtidas através dos escores do preditor, sendo, portanto, noção psicométrica que se aplica à predizibilidade dos escores do teste e não ao conteúdo dos estímulos. Dessa forma, um teste poderá ter numerosas validades para várias situações, o que torna desconfortável a convivência entre as definições conceitual e operacional de validade. Esse é o primeiro problema. Mas há outras questões.

Se nós resolvermos examinar os dois termos principais da definição operacional, penetraremos num paradoxo metodológico. Realmente, fica difícil aceitarmos que se trabalhe o teste com tanta sofisticação estatística para que as inferências obtidas através de seus escores sejam julgadas pela comparação com outra medida elaborada a partir de observações diretas do comportamento, dificilmente submetida a igual nível de tratamento, portanto, sob o ângulo psicométrico, de ordem inferior. É muito comum encontrarmos, na bibliografia especializada, trabalhos minuciosamente elaborados, empregando as mais avançadas técnicas estatísticas, mas que desembocam em rudimentares desenvolvimentos de critérios. Sob vários aspectos, isso é um contra-senso.

Entretanto, podemos supor, pelo menos teoricamente, que o critério receba nível de tratamento estatístico tão alto quanto o teste. Neste caso, poderíamos determinar a validade do teste em função de sua correlação com o critério? O problema que se coloca, dentro dessa linha de pensamento, é que precisaríamos estabelecer, de antemão, a validade do critério, o que implicaria uma cadeia de sucessivas e intermináveis validações de critérios contra

critérios; à impossibilidade prática dessa operação se acrescentaria uma crescente incorreção teórica em face do significado estatístico do coeficiente de correlação.

Mas existe a possibilidade de limitarmos nosso estudo a determinado nível de critério, se ele foi estabelecido pela organização patrocinadora, a partir do diagnóstico organizacional por ela realizado; ou seja, o problema ficaria colocado da seguinte forma: a organização assume a relevância do critério para os seus objetivos, cabendo ao psicólogo, apenas, operacionalizá-lo. Nessa situação, o teste poderia ser validado em função do critério? Suponhamos que o critério estabelecido pela organização para um setor de produção da fábrica seja o número de peças a serem confeccionadas. Se o psicólogo utilizar um teste de percepção de formas como parte da bateria de seleção dos candidatos àquele setor, será lógico que ele verifique a validade desse teste através da quantidade de peças fabricadas pelo pessoal selecionado? Se a validade for considerada, apenas, em termos psicométricos de predizibilidade, isso é até possível; seria uma das inúmeras validades do teste, conforme indicamos anteriormente; porém ela não estaria compatível com a definição conceitual. Mas, se pretendermos validar o teste julgando o seu valor em medir aquilo a que se propõe, ou seja, percepção de formas, através da quantidade de peças produzidas, estaremos empregando procedimento completamente ilógico, mesmo que se considerasse essa característica importante para o desempenho de quem trabalhe em confecção de peças. Um barômetro presta ajuda nas previsões do tempo, mas não seria razoável julgar suas qualidades de barômetro pela precisão daquelas predições (Ebel, 1961). Robert Ebel (1961) considera "um dos maiores escândalos da psicometria" a facilidade com que se induz, os que desenvolvem testes, a aceitarem como medidas de critério dados quantitativos que têm remota e aparente relação com o traço que se quer medir". De qualquer forma, é preciso entendermos que a validade de um teste obtida pela correlação com o critério não pode indicar se ele tem valor para medir aquilo a que se propõe e que geralmente está indicado pelo seu próprio nome, mas apenas se ele é útil para estimar outra medida.

6. Os quatro tipos de validade

Em 1954, a Associação Norte-Americana de Psicologia, procurando esclarecer o conceito de validade utilizado nos testes, publicou os Padrões para Testes Educacionais e Psicológicos, definindo, nessa ocasião, a existência de quatro tipos de validade: preditiva, concorrente, de constructo e de conteúdo; a partir daí, psicólogos de vários outros países passaram a adotar essa classificação. A revisão dos Padrões, publicada em 1974, manteve a mesma classificação, apesar de esses conceitos não terem tanta coisa em comum que a justificasse, pois os dois primeiros baseiam-se na correlação teste-critério, embora realizada de forma diferente, e os dois outros dirigem-se ao conteúdo, focalizando-o, porém, dentro de filosofias que até se opõem. Em nossos comentários anteriores, já mostramos a impropriedade de definirmos operacionalmente a validade sob a forma de correlação teste-critério, ou seja, nos termos das validades preditiva e concorrente. Julgo que será útil examinarmos agora os conceitos de validade de constructo e de validade de conteúdo.

Validade de constructo foi uma expressão criada para indicar o processo de análise do teste, envolvendo a sua inclusão em sistema teórico geral formado por conjunto de regras, supostos e hipóteses que precisariam ser comprovados empiricamente antes que ele fosse declarado possuidor dessa validade (Cronbach & Meehl, 1955). Esse tipo de validade se reporta a atributos postulados ou hipotéticos acerca das pessoas, supondo, muitos psicólogos, que eles se refletem na realização dos testes. Como a validade de constructo não pode ser expressa através de algum padrão de medida, pode dar origem a muita especulação e pouca ciência. Mesmo para a pesquisa teórica, essa forma de validação não parece ter introduzido acréscimo ao conhecido processo científico de construção e comprovação de teorias psicológicas. De qualquer forma, o certo é que, se depois de todas as operações recomendadas para a validação de constructo, quisermos aplicar as suas conclusões à seleção de pessoal, teremos, sempre, que testar uma nova hipótese: a do relacionamento do teste com o desempenho no trabalho; isto é, voltamos ao ponto de partida, porque a validade de constructo, nesse sentido, preocupa-se, apenas, com o caráter intrínseco da medida. Então, para aqueles que trabalham em seleção de pessoal com os comportamentos observáveis, rejeitando atributos hipotéticos, postulados ou latentes, o conceito é desnecessário. Para os que acreditam na existência desses atributos por trás da realização nos testes, a validação de constructo pode ter utilidade no sentido de maior compreensão do que foi postulado. Bechtoldt chegou a declarar que não entendia por que foi criada essa categoria de validade; e Ebel colocou pergunta difícil de responder: "Algum teste educacional ou de emprego já demonstrou possuir validade de constructo?"

O último conceito definido pelos Padrões é o de validade de conteúdo. Novamente o termo validade não é adequado aos testes dessa categoria, pois ela não se refere ao processo de validação psicométrica, mas à técnica de construção, orientada no sentido de torná-los representativos de um domínio de conteúdo. O fundamento para a utilização dos testes baseados na validade de conteúdo, em seleção, está em que: se um teste é amostra representativa da população de conteúdos da tarefa, está logicamente com ela relacionado, como qualquer amostra com a população que representa; em consequência, formula-se a hipótese de que se o candidato realizar o trabalho na situação de teste, também o faça na situação de desempenho do cargo. Como o relacionamento com as tarefas do cargo se apóia no cuidado com que se procurou construir o teste, de modo a torná-lo amostra representativa do trabalho, o conceito de validade aqui utilizado não tem o mesmo significado da validade correlacional, podendo deixar de existir qualquer preocupação com o processo de validação. Robert Ebel demonstrou que a necessidade de ser determinada a validade de um teste cresce na medida em que o comportamento do teste fica mais distante no tempo ou na tarefa em relação ao comportamento do critério; caso esse relacionamento seja muito estreito, deixa de existir interesse pela validade e sim pela fidedignidade ou generalidade; lembra que na própria estatística mede-se a variabilidade de um conjunto de escores, calculando o seu desvio-padrão e ninguém se preocupa com a validade dessa medida. A sua conclusão é que, se o próprio processo de medição define quantitativamente o que vai ser medido, não há necessidade de levantar a questão da validade. Mas então, como saber se um teste possui essa classe de validade?

Os questionamentos que possam surgir em relação à validade de conteúdo devem dirigir-se aos procedimentos empregados na construção do teste, que envolvem, geralmente: a) definição do domínio de conteúdo, em termos de tarefas, atividades etc.; b) coleta de aspectos representativos desse domínio e julgamento dessa representatividade, através de juízes-peritos qualificados; c) avaliação da fidedignidade do julgamento dos juízes.

Alguns psicólogos associam de tal forma o nível científico de um trabalho à tecnologia estatística nele embutida que podem considerar a validade de conteúdo de nível inferior à validade correlacional. Devemos lembrar que as medidas de critérios, a partir das quais são determinadas as validades correlacionais dos testes, constituem, normalmente, medidas baseadas na validade de conteúdo. Talvez Robert Ebel (1977) tenha razão quando diz que “a validade de conteúdo representa o único fundamento básico para qualquer espécie de validade”.

7. Conclusão

Procuramos, ao longo destas considerações, realizar uma análise dos principais conceitos e procedimentos que, tradicionalmente, têm orientado o processo de seleção de pessoal, mostrando quão absurdo é continuarmos a utilizá-los da forma como tem sido feita.

Ao final desta exposição, parece-nos importante reforçar alguns aspectos críticos de nossa posição, ressaltando as seguintes colocações:

1. Até o presente momento não existe base lógica ou científica para atribuímos a categorias descritivas de comportamentos observáveis — como as atuais aptidões e os traços de personalidade — existência própria, independente dos testes que lhes deram origem. A abordagem que temos feito em psicologia através dos testes não permite, de forma alguma, avançarmos conclusões sobre os processos que atuam no indivíduo, enquanto ele procura soluções para as questões que lhe são apresentadas. Dessa forma, não há razões plausíveis para conservar o significado que esses conceitos adquirem em psicologia do trabalho.

2. Se a linha de pensamento seguida pelo psicólogo o levar a presumir existirem atributos latentes que se refletem na resolução das questões do teste e no desempenho das tarefas do trabalho, deverá obrigatoriamente comprovar essa hipótese, porém, para cada teste, *antes* de utilizá-lo na tomada de decisões em relação às pessoas. Essa comprovação deverá obedecer a todas as exigências do modelo científico de pesquisa em seleção.

3. Qualquer que seja a orientação do psicólogo em seleção de pessoal, os testes empregados devem ser desenvolvidos ou escolhidos após terem sido elaborados os critérios e a partir deles.

4. Considerando a validade de um teste como a sua capacidade de medir aquilo a que se propõe, constitui erro pensarmos que se pode validar o teste através de sua correlação com o critério. Entretanto, o estabelecimento de relações teste-critério serve para avaliar a predizibilidade dos escores do teste

para várias situações específicas. De qualquer modo, o termo validação não exprime o que se tem feito com os testes na pesquisa em seleção.

5. Os testes baseados na amostragem de conteúdo constituem, geralmente, a solução mais adequada à seleção de pessoal, sempre que puderem ser desenvolvidos. A qualidade desse tipo de teste vai depender da forma como foi realizada a análise do trabalho, da identificação correta do domínio de conteúdo e do cuidado na sua elaboração, objetivando torná-lo representativo desse conteúdo.

6. Um teste, classificado como de aptidão, tende a relacionar-se mais fortemente com o critério, quando os conteúdos das duas medidas se tornam semelhantes, isto é, quando é conservada tanto quanto possível “a integridade da relação sujeito-objeto”, ou seja, quando ele tende a se tornar um teste de amostragem de conteúdo.

7. Ser avaliado de acordo com a sua real capacidade para exercer o cargo é um direito do indivíduo, um dever da organização e uma necessidade da sociedade, exista ou não legislação reguladora a esse respeito. Dessa forma, procedimentos corretos de seleção de pessoal e direito ao emprego, de acordo com a capacidade real de exercê-lo, devem ser metas interdependentes. Dentro dessa idéia, parece aconselhável que, no Brasil, a nossa profissão se anteceda às restrições legais que deverão surgir em relação aos testes psicológicos, a partir dos primeiros recursos à justiça — a seleção de motoristas, por exemplo, tem todas as condições para abrir o caminho — elaborando normas rigorosas relativas ao uso dos testes psicológicos em seleção de pessoal, mantendo, desse modo, o controle da situação. Não devemos esperar para ver o que vai acontecer.

Abstract

This paper discusses conceptions and procedures that support bases for the use of psychological tests in personnel selection. Data of researches employing occupational tests reveal inexpressive results. Interpretations given to psychological characteristics lack truthfulness. Emphasis is put in the need to prove the relation test-performance before the predictor be used for decision ends. The unfitnes of several concepts of validity is analysed. Citations of restrictions to the use of tests taken by both American government as courts of justice show the need for Brazilian psychologists to propose rigid measures to the use of tests, anticipating the legal controls that will inevitably appear.

Referências bibliográficas

- Bechtoldt, H. P. Construct validity: a critique. *American Psychologist*, 14:619-29, 1959.
- Campion, J. E. Work sampling for personnel selection. *Journal of Applied Psychology*, 56(1): 40-4, 1972.
- Cronbach, L. J. & Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin*, 52:281-302, 1955.
- Dunnette, Marvin D. Aptitudes, abilities and skills. In: ———, ed. *Handbook of industrial and organizational psychology*. Chicago, Rand McNally College, 1976. cap. 11.
- Ebel, R. L. Must all tests be valid? *American Psychologist*, 16:640-7, 1961.

- Ebel, R. L. Comments on some problems of employment testing. *Personnel Psychology* 30:55-63, 1977.
- Faverge, J. M. Quelques aspects rétrospectifs et prospectifs du testage. *Bulletin du CERP*, 1(2):57-65, 1968.
- . *L'examen du personnel et l'emploi des testes*. Paris, Presses Universitaires de France, 1972.
- Ghiselli, E. E. *The validity of occupational aptitude tests*. New York/London/Sydney, John Wiley, 1966.
- . The validity of aptitude tests in personnel selection. *Personnel Psychology*, 26:461-77, 1973.
- Guion, Robert M. Recruiting, selection, and job placement. In: Dunnette, Marvin D., ed. *Handbook of industrial and organizational psychology*. Chicago, Rand McNally College, 1976. cap. 18.
- . Content validity in moderation. *Personnel Psychology*, 31:205-13, 1978.
- & Gottier, R. F. Validity of personality measures in personnel selection. *Personnel Psychology*, 18:135-64, 1965.
- Hough, L. M.; Keyes, M. A. & Dunnette, M. D. An evaluation of three "alternative" selection procedures. *Personnel Psychology*, 36:261-76, 1983.
- Lawshe, C. H. A quantitative approach to content validity. *Personnel Psychology*, 28:563-75, 1975.
- Montmollin, M. *A psicotécnica na berlinda*. Agir, 1974.
- Resnick, L. B. e D. P. Test in America: the current challenge. *International Review of Applied Psychology*, 31(1):76-90, 1982.
- Ruch, F. L. The impact on employment procedures of the Supreme Court decision in the Duke Power case. *Personnel Journal*, p. 777-83, Oct. 1977. (tradução).
- Tenopyr, Mary L. Content-construct confusion. *Personnel Psychology*, 30(1):47-54, 1977.
- Thompson, D. E. & Thompson, T. A. Court standards for job analysis in test validation. *Personnel Psychology*, 35:865-74, 1982.
- Wanderley, W. M. Elaboração de um sistema de regras para análise e construção de testes de aptidão. (parte I — considerações teóricas). *Arquivos Brasileiros de Psicologia Aplicada*, 28(2):28-67, 1976.

