

Temo encerrar estas considerações meio com jeito — ou sem jeito — de «narodnik», de populista russo do tempo dos czares, anterior à revolução bolchevista. Ou com ar de pregador, clamando pela volta, pela conversão a valores tropicais e populares, onde não chega a maquilagem amarela das recentes luminárias da cidade.

Os testes como instrumento de seleção e orientação vocacional

Rubem Eduardo da Silva

Qualidade fundamental dos instrumentos de seleção e orientação vocacional.

Um aspecto característico dos instrumentos usados em Psicologia e Educação é a necessidade de demonstrar em que proporção desempenham o papel que deles é esperado. Não se pode concluir o que medem os testes simplesmente analisando os nomes que lhes são atribuídos. Os testes não medem diretamente aquilo em que estamos interessados. Do comportamento externo constatado pelo teste, são feitas as deduções a respeito das características do indivíduo, objeto de estudo. É por isso que a primeira e mais importante pergunta que deve ser respondida sobre um teste é para que constatação é ele um instrumento válido. É preciso verificar em cada situação, se o teste está medindo aquilo que se deseja que ele meça.

A validade de um teste é um índice de sua qualidade para fornecer uma medida de algo diferente do que ele mede diretamente. Assim, é válido um teste de raciocínio abstrato, quando se pode constatar que um resultado obtido por um indivíduo neste teste indica sua posição com relação a seus companheiros, quanto ao raciocínio abstrato e não apenas quanto ao número de questões que conseguiu acertar. Para se constatar o que mede realmente um teste, será necessária uma série de averiguações que conduzirão à sua validação.

O processo de validação de um teste, que deve ser contínuo, vai possibilitar se aprender cada vez mais sobre o que significam os seus escores, para que sejam interpretados adequadamente. O significado de um teste, segundo Dunnette¹, deve ser deduzido do acúmulo de informações sobre ele. Validação de um teste é um processo semelhante ao da avaliação de uma teoria científica, sendo necessárias diversas observações e suas interpre-

tações lógicas². É preciso encontrar os meios para se investigar até que ponto o desempenho no teste é indicador da qualidade do comportamento em que se está diretamente interessado.

Os tipos de validade dos testes

Os testes podem ser usados para diversos objetivos. Como sua validade indica estão desempenhando os objetivos visados, serão necessários diferentes tipos de investigação para se estabelecer a validade de um teste. Um comitê constituído pela Associação Americana de Psicólogos, Associação Americana de Pesquisas Educacionais e Conselho para Mensuração em Educação dos Estados Unidos, ao estabelecer os «*Padrões para Testes Educacionais e Psicológicos e seus Manuais*» distinguiu três tipos de validade: de conteúdo, relacionada com um critério e de conceito. Pela validade de conteúdo se demonstra se o conteúdo do teste é uma amostra adequada das situações ou da matéria sobre que devem ser tiradas conclusões. A validade relacionada com um critério é obtida comparando-se os escores de um teste com outras variáveis consideradas indicadoras das características ou do comportamento em questão. Este tipo de validade pode ser preditivo, quando o teste é usado para antecipar o critério, um tipo de comportamento com o qual será comparado no futuro; se porém, o critério já existe quando é aplicado, ter-se-á validade simultânea. A validade de conceito é avaliada investigando-se que qualidades um teste mede, isto é, determinando-se até que ponto determinados conceitos podem ser deduzidos do desempenho no teste³.

Os três tipos de validade são apenas conceitualmente independentes. Na realidade eles se interrelacionam. Quando se analisa a validade relacionada com um critério, ao se estudar a capacidade de um teste para predizer um comportamento futuro, por exemplo, pode-se, ao mesmo tempo, analisar seu conteúdo e contribuir com esta análise para se esclarecer a validade de conceito. O mesmo interrelacionamento poderá ser constatado se se tomar como ponto de partida a validade de conceito ou a validade de conteúdo. De acordo com os objetivos visados pelos testes,

poderá ser dada ênfase a determinado tipo de validade, porque se faz necessária tal evidência. Conseqüentemente, não é o teste que determina o tipo de validade, mas o uso que dele se faz. Um estudo completo sobre um teste exigirá, porém, informações sobre os três tipos de validade.

Quando o teste é usado para seleção ou orientação, o que se espera dele é que seja capaz de predizer características do comportamento que deverão atuar no futuro, na escola ou no trabalho. Aqui o teste deverá predizer o comportamento do indivíduo em situações específicas. O desempenho no teste será comparado com um critério, uma medida independente e direta do comportamento que o teste deve predizer. Com este tipo de validade, ter-se-á uma medida da correlação entre duas amostras do comportamento. Como afirma Anastasi⁴, não existe diferença básica entre teste e critério. Ambos são diferentes amostras de comportamento, cuja interrelação permite predizer um com o auxílio do outro. O critério, em si, não é essencialmente superior ao teste, nem deve ser considerado como uma classe de fenômeno à parte. Porque um teste é considerado válido enquanto se relaciona com um critério, haverá então tantas validades para um teste quantos forem os critérios com os quais possa ser relacionado. Não se terá, portanto, a validade de um teste mas as validades que darão margem às diversas interpretações de acordo com as diferentes situações.

Em seleção e orientação, geralmente, por motivo de ordem prática, o teste é submetido a uma validação simultânea para ser usado posteriormente com valor preditivo. A grande vantagem de um teste com valor preditivo de um comportamento futuro é que tanto os indivíduos poderão ter alguma segurança de seu provável sucesso em determinado curso ou atividade, como também as instituições poderão escolher os indivíduos que têm possibilidade de melhor rendimento. O processo de validade simultânea é muito mais simples, prático e econômico do que o da validação preditiva. O problema fundamental é saber se é possível usar, para predizer o comportamento, com certo intervalo temporal, os dados obtidos em uma validação simultânea. O maior obstáculo a essa utilização será a mudança das características das habilidades testadas em virtude do exercício do desem-

penho da atividade que se quer prever. Seria necessário fazer correlação das habilidades testadas antes e depois do desempenho considerado como critério. Como observa Horst⁵, se a função medida pelos instrumentos psicológicos se relaciona significativamente com o desempenho considerado como critério, então o treino neste desempenho, como consequência das atividades escolares ou do trabalho, poderia aumentar o nível geral dos escores, de tal modo que não representariam os resultados que teriam sido obtidos antes da participação na atividade que se quer prever.

A preocupação com os aspectos práticos dos resultados dos testes não deve levar a se descuidar da sua fundamentação teórica. A teoria deve estar sempre presente, como base e força propulsora da prática, nas atividades profissionais. Conforme constata Cattell e Butcher⁶, a teoria psicológica e a pesquisa básica, geralmente, se preocupam mais em saber-se o teste representa adequadamente um conceito, enquanto a pesquisa aplicada é mais freqüentemente interessada na sua capacidade para prever um determinado desempenho. Mas, é imprescindível se entender a razão das correlações encontradas, para que possam ser melhor interpretadas, nas situações concretas, e, melhor proveito delas se possa tirar na prática. Não se pode, portanto, esquecer a validade de conceito quando se tem interesse em validade relacionada com um critério.

Interpretação do coeficiente de correlação

Não se pode analisar o valor dos testes em seleção e orientação sem se referir à interpretação do coeficiente de correlação. Foi graças às técnicas desenvolvidas pela análise do coeficiente de correlação que puderam ser desenvolvidos os estudos que possibilitarão prever o comportamento com base em dados estatísticos. Para que se possa prever determinado tipo de comportamento é necessário se saber até que ponto variações em dois modos de agir estariam relacionadas. O coeficiente de correlação vai informar o grau de relação existente entre dois tipos de comportamento de tal modo que, conhecendo-se as variações de

um, se possa concluir sobre as variações do outro. O coeficiente de correlação, na opinião de Guilford⁷, mais do que qualquer outro elemento estatístico, abriu novos horizontes em Psicologia e Educação, porque possibilitou descobrir e verificar como os elementos e fenômenos estão correlacionados, o que é fundamental para o progresso de uma ciência.

Para se interpretar a correlação existente entre dois fenômenos, pode-se proceder de várias maneiras. Comparando-se os resultados das variáveis independentes — testes — com os da variável dependente — critério — pode-se calcular a possibilidade que tem um indivíduo de obter determinado grau de sucesso a partir de seus escores nos testes, com o auxílio de uma tabela de esperança. Os estudantes que precisam tomar decisão sobre sua escolha profissional poderão ser informados sobre os vários graus de probabilidade de obter sucesso em determinadas áreas.

O índice mais usado para se interpretar a correlação existente entre dois fenômenos é o coeficiente de correlação. Elevando-se este ao quadrado, obtém-se o «coeficiente de determinação». Este indica qual proporção da total variância do critério está associada como a variância da variável preditiva. E, subtraindo-se este resultado da unidade, obtém-se informação sobre a percentagem da variância devida a outras causas. Extraído-se a raiz quadrada deste último resultado, i.e., a unidade menos o quadrado do coeficiente de correlação, obtém-se o «coeficiente de alienação» ($K = \sqrt{1 - r^2}$). Este coeficiente indica a proporção de erro presente na predição com a utilização do teste, quando o critério é expresso em unidade padrão. Se este coeficiente for subtraído da unidade, obtém-se o «coeficiente de eficiência preditiva» ($1 - \sqrt{1 - r^2}$) que indica a percentagem de erro reduzida, com a utilização do teste, do total de erro de uma predição baseada apenas no acaso. Finalmente, um outro índice utilizado para interpretar a correlação existente entre dois fenômenos é o «erro padrão de estimativa» ($S_{est} = DP_c \sqrt{1 - r^2}$) que é, simplesmente, o coeficiente de alienação multiplicado pelo desvio padrão do critério, e deve ser utilizado quando as medidas deste não são apresentadas em unidades padrão.

Analisando os índices acima apresentados, observa-se que o «coeficiente de alienação» é equivalente à percentagem de va-

riância devida a outras causas ($1 - r^2$). O «coeficiente de alienação» e o «erro padrão de estimativa» indicam o tamanho do erro de mensuração. Estes índices são importantes quando, tratando-se de decisões individuais, se deseja saber qual a probabilidade de se obter determinado resultado no critério, sabendo-se seu escore no teste. Levandõ-se em consideração o erro de mensuração, pode-se prever a flutuação possível no resultado obtido e a probabilidade de erro ao se predizer determinado nível da variável dependente — o critério. Se a finalidade principal dos testes fosse predizer resultados exatos a serem obtidos pelos indivíduos, a perspectiva seria muito pessimista para o uso de tais instrumentos, conforme observa Anastasi⁸, uma vez que, mesmo com coeficiente muito elevado, como 0,80, dificilmente encontrado, ainda seria bastante acentuada a possibilidade de erro nos escores preditos. Seria então extremamente raro um teste, realmente, eficiente para predizer um critério. Na realidade, porém, os testes não são utilizados para predizer uma posição exata dos indivíduos, no futuro, mas para predizer que os indivíduos ultrapassarão os resultados considerados críticos na variável dependente. Num processo de orientação vocacional, por exemplo, não há necessidade de se informar ao orientando que probabilidade existe para ele obter tais notas, seu interesse é saber da sua possibilidade de ingressar no curso e obter resultados satisfatórios.

Condições que afetam o coeficiente de validade

O tamanho da amostra estudada afeta diretamente o coeficiente de validade. Com amostras pequenas, será necessário um coeficiente elevado para que seja estatisticamente significativa, podendo, por outro lado, apresentar-se estatisticamente significativa um coeficiente muito baixo, pelo simples fato de ter sido obtido em uma amostra muito grande.

A homogeneidade ou heterogeneidade da amostra afetará também o coeficiente de validade. Com amostras heterogêneas, podem ser obtidos, mais facilmente coeficientes mais elevados do que com amostras homogêneas. Por conta desta sensibilidade do coeficiente de validade à homogeneidade ou heterogeneidade da amostra, alguns autores⁹ advogam sua substituição pelo erro pa-

drão de mensuração, que é independente da variação das características medidas.

Para se chegar a uma conclusão sobre a validade de um teste para selecionar melhor os candidatos a um curso ou trabalho, seria necessário aplicá-lo a todos os candidatos, admiti-los e depois relacionar os resultados obtidos no desempenho das atividades para que foram escolhidos (critério) com os do teste. É evidente que este processo é por demais dispendioso e não pode ser executado numa situação prática. O que se faz, na realidade, é relacionar os resultados dos candidatos já selecionados. A homogeneidade provocada pela seleção reduzirá então o coeficiente de validade. O problema estaria resolvido se, ao invés do coeficiente de validade, fosse utilizado o erro padrão de mensuração. Este índice, porém, tem suas limitações, pois, sendo expresso em termos de unidades de escala do critério, só pode ser utilizado para comparar a eficiência de duas ou mais situações de predição se as escalas do critério forem comparáveis. O coeficiente de correlação, por ser um número abstrato, independente das unidades de mensuração, pode ser utilizado como base para comparar a acuidade das várias situações de predição, mesmo que unidades diferentes e não comparáveis estejam envolvidas.

Coeficiente de validade e valor do teste

Ao se analisar a contribuição dos testes usados em seleção e orientação, tomando-se como ponto de referência o «coeficiente de eficiência preditiva», o coeficiente de validade deveria ser muito elevado para que o uso de tais testes se revelasse de alguma utilidade. Um teste com um coeficiente de validade de 0,80 contribuiria para uma redução de apenas 40% do erro provocado por uma predição baseada só no caso. Apoiando-se nesta interpretação é que Hull¹⁰, na década de 20, exigia um coeficiente de validade em torno de 0,45 para que pudesse ser considerado de alguma utilidade prática. Só com um coeficiente deste nível se começa a ter 10% de redução do erro devido a uma decisão com base apenas no «palpite». Cronbach e Gleser¹¹, em sua obra pioneira, que se tornou clássica, sobre teoria da decisão,

demonstrou que o «coeficiente de eficiência preditiva» não é um índice adequado para descrever o valor do teste na maioria das situações em que devem ser tomadas decisões.

Ao se considerar o valor informativo do teste, a atenção deve ser dirigida para o que é acrescentado aos elementos existentes, uma vez que já são úteis para uma tomada de decisão. Não se pode considerar a contribuição trazida pelo teste, tendo como ponto de partida uma decisão baseada apenas no acaso, porque, como observa Cronbach¹², na maioria das situações em que o teste pode ser usado, já existe uma quantidade substancial de informações, que possibilitam uma base para uma decisão melhor do que a baseada no simples acaso.

A contribuição de um teste com determinado coeficiente de validade, num processo de seleção, é relativa, porque depende, entre outros, de dois elementos: a) percentagem de candidatos selecionados; b) percentagem dos candidatos bem sucedidos, posteriormente, que teriam sido escolhidos sem o auxílio do teste. Taylor e Russell¹³, elaboraram tabela para mostrar como varia a contribuição do teste em virtude do que denominaram de «razão de seleção» e «razão de sucesso».

Anastasi¹⁴, comentando uma das tabelas de Taylor e Russell, tomando por base a razão do sucesso de 60%, sem uso de teste, mostra como, mesmo com o coeficiente de validade máximo (1,00), ganha-se apenas 3% quando a razão de seleção é de 95%. Por outro lado, quando a razão de seleção é de apenas 5%, a contribuição de um teste com validade de 0,30 pode ser de 22%. Da mesma maneira, quando é por demais elevada a «razão de sucesso» sem uso dos testes, será insignificante a contribuição destes, por mais elevado que seja o coeficiente de validade, podendo ser substancial, porém, no caso de ser muito baixa a «razão de sucesso» sem uso de testes.

A experiência tem demonstrado, conforme constata Guilford¹⁵, que o coeficiente de validade dos testes tem variado entre 0,00 e 0,60, com a predominância na parte mais baixa da distribuição. Os psicólogos e educadores já começam a abandonar a idéia de insistir em coeficientes elevados como 0,70 e

0,80. Não resta dúvida que seria muito mais interessante se trabalhar com testes de elevado coeficiente de validade. Consciente da sua responsabilidade em se empenhar para obter sempre mais elevado nível de coeficiente de validade nos seus testes, o especialista nesta área não pode deixar de constatar que a experiência de mais de meio século tem demonstrado que não se deve esperar dos testes, padrões de validade considerados ideais. Não se pode deixar de reconhecer que testes com coeficientes de validade muito baixo, entre 0,30 e 0,20, podem dar uma real contribuição, justificando sua inclusão numa bateria, sobretudo se medem algo diferente dos outros.

Predição utilizando grupo de testes

O uso de mensurações múltiplas em Psicologia e Educação, que tem sua base na teoria dos traços e fatores, pode ter em vista, conforme Mosier¹⁶, três finalidades distintas. Em primeiro lugar, várias observações devem ser feitas se se deseja obter graus refinados de discriminação entre os indivíduos. Se, por exemplo, é aplicado um teste de cinco questões do tipo tudo ou nada, com um ponto para resposta certa e zero para resposta errada, apenas seis distinções podem ser feitas entre os indivíduos. Se um maior número de diferenciações é desejado, maior deverá ser o número em questões. Pressupõe-se aqui precisão absoluta nas observações feitas, o que é impossível na prática.

Surge então a segunda finalidade importante no uso da mensuração múltipla: elevar o índice de precisão. Sabendo-se que cada observação contém alguma imprecisão, aumentando-se as observações da mesma característica, será obtido um resultado composto, com um índice mais elevado de precisão. Para que isto aconteça, porém, é imprescindível que as diversas medidas sejam interrelacionadas, uma vez que deverão representar a mesma característica fundamental. O que se pretende com os vários testes é diminuir o erro aleatório, assim como os efeitos sistemáticos de outros fatores não relacionados.

Finalmente, o terceiro objetivo da mensuração múltipla é proporcionar informações sobre aspectos não relacionados do

comportamento a ser predito. Este é o objetivo visado quando vários testes são agrupados para formar uma bateria quando se tem em vista medir vários aspectos de aptidão para prever o sucesso na escola. Aqui, a finalidade não é repetir várias observações do mesmo aspecto do comportamento mas, medir diferentes aspectos do critério. Devido à complexidade do critério, vários elementos do comportamento devem ser considerados. Devem ser escolhidos para fazer parte da bateria, os testes que apresentarem índices mais elevados de correlação com o critério e menores entre si.

Para ressaltar a diferença entre o segundo e o terceiro objetivos acima, imaginem-se duas situações em que os dois objetivos possam ser considerados. Suponha-se que três testes 1, 2 e 3 devam ser escolhidos para prever determinado critério. Se este é simples, dependendo de um determinado fator «A», os testes deverão medir este mesmo fator de tal modo que, na hipótese de cada um medir além do fator comum «A» um fator específico, a contribuição da variância dos testes 1, 2 e 3 seja maior do que a de cada fator isolado. A soma dos escores dos três testes será, portanto, menos afetada pelos fatores específicos e muito mais representativa do fator «A» do que os escores de cada um dos três testes. Se, porém, o critério é complexo, relacionado com três fatores, por exemplo, A, B e C, os testes deverão ser escolhidos de tal modo que cada um meça um fator específico e não haja duplicação. O desejado então é que os escores de cada um dos testes dependa de um fator separado e de fatores próprios a cada um deles de tal modo que a interrelação entre os três testes seja zero.

Correlação múltipla e equação de regressão

Com vários testes para prever um critério complexo, o problema que se enfrenta é o de se chegar à melhor maneira de agrupá-los para se obter o máximo rendimento com eles. Sabendo-se que a contribuição de cada teste depende, em proporção direta, da sua correlação com o critério, e, inversa, desta correlação com os outros testes, na composição do escore total os pesos

a serem dados aos testes devem levar em conta estas duas variáveis, de tal modo que as discrepâncias entre o escore composto e a medida de sucesso, a ser predita, sejam as menores possíveis. A técnica adotada para se conseguir tal objetivo tem sido a correlação múltipla, que proporciona os pesos para a equação de regressão múltipla. Os pesos aí obtidos são os mais satisfatórios se o critério adotado é o menor quadrado da diferença entre o sucesso predito e a medida de sucesso obtido. Através deste método, denominado dos mínimos quadrados, os pesos são escolhidos de tal modo que a soma das diferenças entre sucesso predito e obtido seja a mínima. A correlação múltipla proporciona então a mais eficiente combinação de testes com seus respectivos pesos para que se possa obter o máximo poder preditivo dos testes usados.

Alguns pressupostos são necessários para que a correlação múltipla possa atingir seus dois objetivos fundamentais: 1) determinar os pesos a serem aplicados a cada teste de tal modo que o total dos quadrados das discrepâncias entre o comportamento predito e o observado para um grupo de indivíduos seja o menor possível; 2) apresentar uma estimativa da acuidade da predição. O pressuposto fundamental para utilização adequada da equação de regressão múltipla, é que o comportamento considerado como critério seja proporcional a uma soma dos escores dos vários testes juntamente com seus pesos. Pressupõe-se então que haja compensação entre as habilidades necessárias para a adequação do comportamento, de tal maneira que o indivíduo alcance um nível satisfatório de desempenho, apesar da deficiência de algumas habilidades necessárias que seriam compensadas pelo alto nível das outras. É pressuposto também que o aumento do nível de uma das habilidades, mantendo-se constantes as outras, aumente sempre o nível do critério.

Para que se obtenha um melhor resultado, com a utilização dos vários testes, para prever, através da correlação múltipla, é necessário que cada um dos testes possa dar sua contribuição para a bateria através da mensuração de um novo fator atuante no critério e ainda não presente nos outros. O que se constata, na prática, porém, é que, conforme observa Cronbach¹⁷, a correlação múltipla atinge rapidamente um platô,

de tal modo que os testes que se acrescentam, depois do terceiro ou do quarto, raramente trazem alguma contribuição de valor. A grande dificuldade é encontrar testes que, incluídos numa bateria, possam dar uma contribuição única ao conjunto dos já existentes, porque, em geral, eles se superpõem na mensuração dos fatores atuantes no critério.

A variável «supressiva»

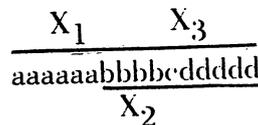
Um fenômeno paradoxal pode ocorrer na correlação múltipla quando, uma variável independente, apesar de sua correlação zero com a dependente, pode contribuir para aumentar o nível de correlação por conta de sua correlação com as outras variáveis independentes. Esta variável é denominada «supressiva» porque, não medindo as variâncias do critério, exerce sua ação suprimindo das outras medidas preditivas as variâncias comuns a ela e não presentes no critério.

McNemar¹⁸ explica, através de um diagrama que produzimos abaixo, como atua a «variável supressiva» para elevar o nível da correlação múltipla. O autor parte da idéia de que uma correlação pode ser interpretada em termos de elementos comuns às duas variáveis e pode ser obtida através da seguinte fórmula:

$$r_{xy} = \frac{n_c}{\sqrt{n_x + n_c} \sqrt{n_y + n_c}}$$

onde n_x é igual ao número de elementos únicos a X; n_y , o número de elementos únicos a Y; e n_c , o número de elementos comuns a X e Y.

Supondo que se tenha um conjunto de três variáveis: X_1 , X_2 e X_3 , as duas primeiras com 10 elementos cada, e a terceira com cinco; e, que X_1 e X_2 têm quatro elementos em comum; X_2 e X_3 , cinco; X_1 e X_3 , nenhum; o seguinte diagrama representaria as três variáveis com interrelação de seus elementos.



Aplicando-se a fórmula acima seriam obtidas as seguintes correlações:

$$r_{12} = 0,400 \text{ , } r_{13} = 0,000 \text{ , } r_{23} = 0,707$$

Considerando a variável X_1 como critério e as variáveis X_2 e X_3 como preditivas, teríamos os seguintes pesos: beta 2 (B_2) = 0,800 e beta 3 (B_3) = - 0,566; e a correlação múltipla seria $R_{1.23} = 0,566$. Apesar da correlação nula da variável X_3 com o critério e o peso negativo, houve um substancial aumento com a utilização desta variável na correlação múltipla. Este resultado se deve ao fato de a variável X_3 suprimir os efeitos dos elementos presentes na variável X_2 , comuns a estas duas variáveis e ausentes do critério. Considerando ausentes os elementos «d», a correlação entre X_1 e X_2 , utilizando a mesma fórmula anteriormente apresentada, seria 0,566, resultado igual ao da correlação múltipla. A variável X_3 funcionou, portanto, para suprimir os efeitos dos elementos não relevantes e assim contribuiu para a correlação múltipla apesar de ser nula sua correlação com o critério.

Embora não seja muito freqüente se encontrar, na prática, o efeito da «variável supressiva» explicado conceitualmente por McNemar, algumas constatações são apresentadas pelos autores. Horst¹⁹ apresenta um caso de atuação da «variável supressiva» na predição de sucesso no treinamento de pilotos durante a segunda guerra mundial. Foram utilizados quatro testes de aptidão na seleção dos candidatos: habilidade mecânica, numérica, espacial e verbal. Os três primeiros apresentaram correlações positivas e significativas com o critério, enquanto o quarto revelou correlação muito baixa com o critério e, razoavelmente elevada com os outros testes. Apesar de ser negativo o peso dos scores do teste de habilidade verbal, a correlação múltipla foi mais alta com a inclusão deste. Analisando-se o fenômeno, constatou-se que, embora a habilidade verbal não fosse importante para o treinamento dos pilotos, era necessária para o bom desempenho nos outros três testes de aptidão, uma vez que possibilitaria compreender as instruções e as questões usadas nestes. Incluindo o teste de habilidade verbal, com peso negativo na

equação de regressão, se suprimiria habilidade irrelevante para o critério, reduzindo-se os escores daqueles que se saíram bem nos testes, mais por conta da habilidade verbal do que das habilidades exigidas para o sucesso no treinamento dos pilotos.

O método do escore múltiplo crítico

Em situações em que algumas habilidades fossem tão importantes que não pudessem ser substituídas por outras, não seria possível então a compensação quando as primeiras habilidades se apresentassem em nível muito abaixo do necessário para o desempenho esperado. Neste caso, não seria possível utilizar a equação de regressão múltipla, sendo então necessário usar o método «escore múltiplo crítico». Este método consiste em determinar escores mínimos para aquelas habilidades essenciais, abaixo dos quais os candidatos não seriam aceitos. Estes escores mínimos deverão ser determinados na base do julgamento, através do processo de ensaio e erro. A necessidade da utilização de escore crítico se apresenta quando se trata de habilidades exigidas para dirigir automóvel, como observa Mosier²⁰. Por mais elevadas que sejam as outras habilidades para dirigir, se a acuidade visual do indivíduo for abaixo de determinado limite não poderá dirigir com segurança.

Em todas as situações em que for possível a sua utilização, o uso da equação de regressão múltipla é preferível ao do método de «escore múltiplo crítico». Este é, sem dúvida, mais simples de ser aplicado quando já foram estabelecidos os escores mínimos. Conforme observa Guilford²¹, arrisca-se mais ao se rejeitar alguém com base em um teste do que no conjunto de vários, porque neste último caso a precisão é certamente maior. Se os testes que fazem parte de uma bateria apresentam alguma intercorrelação espera-se que o total de escores apresente maior precisão do que uma parte deles. Uma outra grande vantagem da equação de regressão múltipla é que permite uma avaliação de cada indivíduo situando a sua posição no grupo em que se encontra, enquanto com o outro método se sabe apenas se os indivíduos foram aceitos ou não, se se situaram ou não em determinado nível. Em muitas situações, como observa Anasta-

si²², a melhor estratégia é utilizar uma combinação dos dois processos. Aplica-se inicialmente o método do «escore múltiplo crítico» para rejeitar os indivíduos abaixo dos escores mínimos e depois se aplica a equação de regressão múltipla para se obter a posição relativa dos indivíduos com relação ao critério.

Seleção e classificação através de testes

Em Psicologia e Educação, o uso dos testes deve oferecer informações para decisão não apenas quanto a um, mas a vários critérios. O problema da decisão a ser tomada tanto pela instituição como pelo indivíduo é, como observam Cronbach e Gleser²³, na sua forma mais ampla, um problema de classificação, que inclui também a orientação vocacional.

A classificação se distingue da seleção porque nesta uma das alternativas é a rejeição do indivíduo. Uma outra distinção a merecer menção entre seleção e classificação é que neste são dois ou mais critérios a serem preditos pelos testes. A decisão vocacional é basicamente uma classificação feita pelo indivíduo após comparar as várias informações com relação aos diversos critérios. A classificação feita por uma instituição tem por objetivo predizer os vários níveis de desempenho nos diversos critérios a serem apresentados pelos diferentes indivíduos. Na orientação vocacional existe apenas um indivíduo diante de uma imensa quantidade de critérios, enquanto na classificação são vários indivíduos com um número mais reduzido de critérios.

As exigências feitas de uma bateria de testes numa classificação são maiores do que numa seleção. Os testes devem ter diferentes validades para os vários critérios. Se os testes apresentarem a mesma validade para os diversos critérios não têm valor para uma classificação porque não poderiam indicar os diferentes níveis de desempenho a serem esperados dos indivíduos. Na classificação, os testes terão que informar não apenas que tais indivíduos têm possibilidade de sucesso em determinadas atividades, mas deverão informar que certos indivíduos poderão desempenhar com maior sucesso atividade A do que atividade B, C, etc.

Para que os testes desempenhem então seu papel, satisfatoriamente, num processo de classificação, deverão, como afirmam Thorndike e Hagen²⁴, no conjunto, apresentar uma validade ampla para todos os critérios e, individualmente, revelar uma «validade diferencial» para cada um dos critérios. O valor de cada um dos testes será proporcional à sua «validade diferencial», a qual, quando se trata de dois critérios, foi expressa por Cronbach²⁵ por meio da seguinte fórmula:

$$s_1 r_{1t} - s_2 r_{2t}$$

sendo s_1 e s_2 os desvios padrão dos dois critérios; r_{1t} e r_{2t} os coeficientes de validade preditiva do teste «t» com os dois critérios.

Uma bateria ideal para classificação seria aquela em que os testes apresentassem correlação positiva elevada com um critério e correlação baixa ou mesmo negativa com os outros. Tal ideal poderia ser obtido com menos dificuldades se os critérios apresentassem correlação muito baixa ou nula entre si.

O método clínico no processo de seleção

As técnicas até aqui discutidas para utilização das informações apresentadas pelos testes com vistas a uma decisão a ser feita sobre os indivíduos ou pelos indivíduos são, simplesmente, aplicações de fórmulas que podem ser feitas mecanicamente. Todos os indivíduos são considerados da mesma maneira, tratando-se objetivamente as informações obtidas sem se preocupar com aspectos característicos próprios de cada indivíduo em cada situação. Poder-se-ia então perguntar se melhores resultados não poderiam ser obtidos através de julgamentos feitos por profissionais experimentados, a partir dos dados obtidos e com acréscimo de informações que não podem ser fornecidas pelos testes. A questão aqui colocada é se o tratamento clínico dos resultados dos testes é mais eficiente do que o tratamento estatístico.

Muitos afirmam que os métodos objetivos no tratamento dos dados para uma seleção deixam de lado uma grande quanti-

dade de informações válidas que poderiam ser utilizadas por um clínico qualificado que julgaria, em cada caso, a adequação dos dados a cada indivíduo. O clínico poderia usar muitos dados que surgem na entrevista, por exemplo, e que não podem ser obtidos através dos instrumentos de mensuração existentes. Não estando preso a um esquema rígido de uma equação de regressão, poderia adaptar-se às situações particulares de cada indivíduo para uma melhor utilização das informações.

É preciso distinguir duas funções diferentes que podem ser desempenhadas pelo método clínico. Em primeiro lugar, muitas informações obtidas através de entrevistas não poderiam ser fornecidas pelos escores dos testes. Tanto as informações obtidas pelo método clínico como as obtidas pelo método estatístico são importantes num processo de seleção, tendo cada tipo de informação seu aspecto peculiar. As informações obtidas pelo método podem ser submetidas a tratamento estatístico para se poder constatar a sua contribuição para o processo de seleção. É difícil dizer qual dos dois métodos é o melhor porque, como observa Anastasi²⁶, cada um desempenha melhor seu papel no seu próprio domínio, sendo a situação ideal, para um, diferente da do outro. Quando os dois métodos são comparados na mesma situação um dos dois poderia ser beneficiado.

Uma segunda função é o tratamento clínico dos escores dos testes ou outros dados num processo de seleção. A indagação aqui é se o clínico pode tomar decisões mais adequadas com base na sua experiência do que através de dados estatísticos, como a equação de regressão. Será que um profissional competente, utilizando a sua experiência, obterá melhor resultado num processo de seleção do que alguém que se basear simplesmente nas fórmulas fornecidas pelo método clínico?

Thorndike²⁷ afirma que a grande vantagem do método clínico na avaliação dos escores é a flexibilidade de tratamento que pode ser dada, não se restringindo a uma combinação linear apenas. Esta flexibilidade, porém, está condicionada a dois pressupostos que muito dificilmente se fazem presentes: 1) que o critério não possa ser predito por uma combinação linear dos escores dos testes; 2) que haja muitos profissionais capazes de

descobrir a melhor combinação dos escores dos testes para melhor predizer o critério. A experiência tem demonstrado que o método clínico para interpretação dos testes não se tem revelado superior ao método estatístico. Meehl²⁸, em 1954, fez uma revisão exaustiva da literatura sobre o problema e dos 20 casos estudados, em que se fazia comparação dos dois métodos, com exceção de um caso apenas, as predições feitas com o método estatístico eram superiores ou aproximadamente iguais às feitas pelo método clínico. Em 50% dos estudos feitos até 1954, o método estatístico se revelou superior ao método clínico. Em 1966, Sawyer²⁹, revendo a literatura sobre o assunto, concluiu que os clínicos não melhoram as predições estatísticas, mesmo quando usam os dados estatísticos como parte de suas informações. Os dados sugerem que a melhor contribuição dada pelos clínicos não é quando eles fazem as avaliações finais a partir de dados, mas quando suas informações são introduzidas com os outros dados, numa equação de regressão ou outro processo de combinação estatística que irá possibilitar avaliação final.

Cronbach³⁰, analisando os resultados obtidos pelo método clínico, conclui que a razão principal da reduzida eficiência deste método é que os clínicos combinam dados por meio de uma maneira intuitiva de dar pesos, cuja adequação não é testada posteriormente. Por outro lado, as fórmulas estatísticas são cuidadosamente verificadas em amostras representativas do universo a que pertencem os indivíduos testados; e são aplicadas sempre da mesma maneira, atribuindo sempre os maiores pesos às variáveis que mais contribuem para predição. Os clínicos, por sua vez, não têm a mesma constância na atribuição dos pesos e são freqüentemente, influenciados por preconceitos e estereótipos, tratando muitas vezes de modo diferente os indivíduos por conta da diferença de sexo, por exemplo, sem que haja evidência para sua maneira de agir.

BIBLIOGRAFIA

- 1 — DUNNETE, Marvin D. *Personnel selection and placement*. Belmont, Wadsworth, 1966, p. 113.
- 2 — CRONBACH, Lee J. Test validation. In: THORNDIKE, Robert L., ed. *Educational measurement*. 2. ed. Washington, American Council on Education, 1971, p.445.

- 3 — STANDARDS for educational and psychological tests and manuals. Washington, American Psychological Association, 1966, p. 12-3.
- 4 — ANASTASI, Anne. The concept of validity in the interpretation of test scores. In: CHASE, Clinton J. & LUDLOW, H. Glenn, ed. *Readings in educational and psychological measurement*. Boston, Houghton Mifflin Co. 1966, p. 64.
- 5 — HORST, Paul, *Psychological measurement and prediction*. Belmont, Books/Cole Pub. Co., 1968, p. 345.
- 6 — CATTELL, Raymond B. & BUTCHER, H. J. *The prediction of achievement and creativity*. New York, The Bolbs-Merrill Co., 1968, p. 92.
- 7 — GUILFORD, J.P. *Fundamental statistics in psychology and education*. 3. ed. New York, McGraw-Hill, 1956 p. 135.
- 8 — ANASTASI, Anne. *Psychological testing*. 3. ed. New York, Macmillan, 1968, p. 131.
- 9 — BLOMMERS, Paul & LINDQUIST, E.F. *Elementary statistical methods in psychology and education*. Boston, Houghton Mifflin Co., 1960, p. 432.
- 10 — HULL, C.L. Aptitude testing. Yonkers, N.Y.: World, 1928 Cap. 8. Apud GUILFORD, J.P. *Fundamental Statistics in psychology and education*. 3. ed. New York, McGraw-Hill 1956. p. 146.
- 11 — CRONBACH, Lee J. & GLESRE, Goldine C. *Psychological tests and personnel decisions*. 2. ed. Urbana, University of Illinois Press, 1965, p. 67.
- 12 — New light on test strategy from decision theory. In ANASTASI, Anne, ed. *Testing problems in perspective*. Washington, American Council on Education, 1966, p.54-5
- 13 — TAYLOR, H.C., & RUSSELL, J.T. The relationship of validity coefficients to the Practical effectiveness of tests in selection. *J. appl. Psychol.*, 1939,23,265-78. Apud GUILFORD, J.P. *Fundamental Statistics in psychology and education*. 3. ed. New York, McGraw-Hill, 1956. p.380.
- 14 — ANASTASI, Anne. Op. cit. p. 134-5.
- 15 — GUILFORD, J.P. Op. cit. p. 146.
- 16 — MOSIER, Charles I. Batteries and profiles. In: LINDQUIST, E.F., ed. *Educational measurement*. Washington, American Council on Education, 1966, p. 770-3.
- 17 — CRONBACH, Lee J. *Essentials of psychological testing*. 2.ed. New York, Harper & Raw, 1960, p. 340.
- 18 — McNEMAR, Quin. *Psychological statistics*. 3. ed. New York, J. Wiley, 1962, p.186-7.
- 19 — HORST, Paul. Op. cit. p. 355
- 20 — MOSIER, Charles I. Op. cit. p. 785.
- 21 — GUILFORD, J.P. Op. cit. p. 429.
- 22 — ANASTASI, Anne. Op. cit. p. 151.
- 23 — CRONBACH, Lee J. Op. cit. p. 12.
- 24 — THORNDIKE, Robert L. & HAGEN, Elizabeth. *Measurement and evaluation in psychology and education*. 3. ed. New York, J. Wiley, 1969, p. 636.

- 25 — CRONBACH, Lee J. Op. cit. 357.
- 26 — ANASTASI, Anne. Op. cit. p. 152.
- 27 — THORNDIKE, Robert L. *Personnel selection. Test and measurement techniques.* New York, J. Wiley, 1966, p. 201.
- 28 — MEEHL, Paul. *Clinical versus statistical prediction. A theoretical analysis and a review of evidence.* Minneapolis, University of Minnesota, 1966, p. 119.
- 29 — SAWYER, J. Measurement and prediction, clinical and statistical. *Psychological Review*, 1966, 66, 178-200. Apud HILLS, John R. Use of measurement in selection and placement. In: THORNDIKE, Robert L., ed. *Educational measurement.* 2. ed. Washington, American Council on Education, 1971, p. 697.
- 30 — CRONBACH, Lee J. Op. cit. 347.

VOCACIONAL