

Técnica geral de organização de provas objetivas

TOMÁS DE VILANOVA MONTEIRO LOPES

Técnico de Administração

"To write a good examination the examiner must know the job backwards and forwards". — MOSHER & KINGSLEY (1).

A VERIFICAÇÃO de conhecimentos é de capital importância na seleção de pessoal para o Serviço Público. Tal verificação, na maioria dos casos, se faz por meio de provas escritas. Uma boa prova escrita para verificação de conhecimentos deve possuir, entre outros, os seguintes predi-

1. adequação ao que se deseja medir;
2. constância na medida;
3. facilidade de aplicação;
4. boa chave de correção.

A prova objetiva está, precisamente, neste caso.

Prova objetiva é aquela que permite eliminar a influência do elemento subjetivo, na avaliação dos resultados (2).

Seu característico mais ostensivo é a imparcialidade. Assegurando um julgamento impessoal e uniforme, a prova objetiva oferece reais possibilidades de atingir a finalidade do instrumento do exame, qual seja a medida das aptidões ou conhecimentos do indivíduo, sem que os mesmos sejam adulterados pelas reações subjetivas do examinador.

Antes de iniciar a construção da prova objetiva devemos procurar obter a especificação:

- a) dos tipos de atividades compreendidos pelo trabalho para cujo exercício os indivíduos vão ser selecionados;
- b) dos conhecimentos necessários para o exercício eficiente dos citados tipos de atividade.

Como obter essa especificação? O melhor recurso de que podemos dispor para obter uma

boa tábua de especificações é a análise do trabalho.

Compreendemos por análise do trabalho a operação através da qual se determina o conteúdo do trabalho, do ponto de vista tecnolológico, e se estabelecem as condições necessárias para o seu exercício.

PRESTON ROBINSON (3) entende que análise do trabalho é o estudo científico de cada fase do trabalho, incluindo a natureza deste e o tipo de trabalhador necessário para executá-lo.

EDWARD STRONG e RICHARD UHRBROCK (4) preferem considerá-la como sendo um método de dissecar cientificamente um trabalho, de modo a determinar seus elementos componentes e a influência dos mesmos sobre a aprendizagem, a produção e a movimentação de pessoal.

Muito embora não se possa estabelecer uma regra estritamente aplicável a todos os casos, tem-se admitido, geralmente, que a análise do trabalho deve fornecer elementos que tornem possível:

1. a identificação do trabalho;
2. a enumeração detalhada e explicativa das atribuições e responsabilidades de quem executa o trabalho;
3. a descrição dos materiais e do equipamento empregados na execução do trabalho;
4. o conhecimento do ambiente em que o trabalho é executado;
5. a determinação das aptidões e dos conhecimentos que caracterizam o trabalhador eficiente.

O êxito da análise do trabalho depende, naturalmente, da propriedade do método empregado. A rigor, podemos dispor de dois métodos:

- a) o do estudo dos tempos e dos movimentos;
- b) o da descrição do trabalho.

O primeiro oferece pouco interesse para nós, não apenas por ser passível de justas e apreciáveis

(1) Os números fazem remissão à bibliografia indicada no final.

restrições do ponto de vista técnico (5), senão também por ser de uso muito limitado nos domínios da seleção de pessoal para os serviços públicos.

Com efeito, o estudo dos tempos e dos movimentos não tem grande utilidade quando o trabalho, como acontece de modo geral no Serviço Público, não é predominantemente manual nem obedece a condições uniformes, estandardizadas, que permitam fazer corresponder às unidades de tempo e às fases de movimentos certos padrões de produção.

Por isso mesmo, devemos, em nosso caso, preferir o método da descrição do trabalho. Este método utiliza, como instrumentos fundamentais:

- a) a experiência pessoal do analista;
- b) a observação (que focaliza o trabalhador no exercício do trabalho);
- c) a informação.

O ideal seria, sem dúvida, o emprego desses três elementos conjugados. Comumente, porem, nas grandes organizações, mesmo quando a análise do trabalho é feita para fins de organização de provas, considera-se suficiente a informação obtida por meio de questionários.

Alguns autores, entre eles SOLLIER e DRABS (6) e FRANZISKA BAUMCARTEN (7), põem em dúvida o valor dos questionários como meio suficiente para a coleta de dados que assegurem uma satisfatória descrição do trabalho.

Mas o certo é que as razões arguidas por esses autores quase que já não subsistem em face dos aperfeiçoamentos introduzidos na técnica de organização dos questionários e nos processos de apuração e tratamento dos dados coligidos.

A crença hoje dominante é a de que os questionários nos permitem, perfeitamente, conhecer os aspectos essenciais e característicos do trabalho a descrever, tudo dependendo de serem observadas na sua organização certas cautelas e uma sincera atitude de pesquisa.

Os processos adotados na prática da descrição do trabalho por meio de questionários são diversos. Um processo que, a nosso ver, proporciona resultados compensadores e, até certo ponto, conjuga os principais instrumentos do método, consiste em:

PRIMEIRA FASE (*estudo*):

1. Estudar previamente o trabalho e suas condições de execução, para isso recorrer, de preferência, às seguintes fontes:
 - a) legislação relativa às atividades do ramo;
 - b) regulamentos de serviço (*idem*);
 - c) regulamentos de cursos profissionais (*idem*);
 - d) monografias profissionais (*idem*);
 - e) livros, revistas e outras publicações (*idem*);
 - f) depoimentos de profissionais qualificados;
 - g) observação pessoal.

2. Enumerar, em face dos dados colhidos, o maior número possível de operações compreendidas pelo trabalho. Afim de facilitar a apuração posterior: a) designar essas operações com os nomes pelos quais elas são, comumente, designadas na gíria dos profissionais do ramo; b) traduzi-las, sempre que possível, em termos de atitude.

3. Enumerar os conhecimentos presumivelmente necessários à execução das operações acima e dividi-los, quanto à natureza (v.g. teóricos, práticos) e quanto ao nível e extensão, observando-se, até onde o permitirem as peculiaridades de cada caso, a divisão oficial do ensino (cursos, séries etc.).

4. Organizar uma tabela provisória, onde figurem, paralelamente, para efeito de fácil confronto, a enumeração das operações compreendidas pelo trabalho e os conhecimentos considerados necessários.

SEGUNDA FASE (*inquérito*):

5. Realizar um inquérito entre os trabalhadores do ramo, organizando-se, para isso, um questionário escrito, cuja orientação geral deverá ser determinada pelo estudo anterior (fase 1.^a). Quanto ao aspecto propriamente formal, convém que o questionário obedeça a umas tantas condições, v.g.:

a) não deve ser muito extenso, afim de poder manter a necessária concentração da atenção e do interesse em torno dos pontos cujo conhecimento é essencial;

b) não deve recorrer a processos de memória forçada nem conter elementos de sugestão capazes de eliminar a espontaneidade das respostas;

c) não deve ser vasado em linguagem imprópria (vaga, ambígua, obscura ou inacessível à mentalidade de quem irá preenchê-lo);

d) deve evitar perguntas cujas respostas envolvam julgamento sobre os aspectos do trabalho diretamente relacionados com os interesses pessoais do trabalhador;

e) deve evitar as enumerações para escolha; as enumerações, quando empregadas, deverão ter caráter meramente explicativo ou exemplificativo;

f) deve ser um só para todos os trabalhadores, de modo a assegurar uma base comum para a apuração e apreciação dos dados;

g) deve, em seu conjunto, ser decalcado numa concepção clara dos objetivos em vista e limitar-se a pontos bem determinados.

TERCEIRA FASE (apuração e interpretação):

6. Proceder à apuração e tabulação dos dados coligidos no inquérito e realizar o trabalho de interpretação.

7. Preparar a tábua de especificações, à base da interpretação dos resultados do estudo e do inquérito, classificando-se os conhecimentos e os tipos de trabalho, pela frequência de utilização e de realização, em *indispensáveis, complementares e menos importantes*.

PARTE II

"One of the major defects of typical examinations has been the fact that have given evidence only to a limited number of objectives and have not adequately indicated the degree to which students were attaining all of the desired outcomes of instruction".

— HAWKES, LINDQUIST & MANN. (8).

Uma vez que a tábua de especificações já foi organizada e, por consequência, já foram determinados os tipos de trabalho e os conhecimentos necessários para a sua execução, a próxima etapa será a da construção da prova de ensaio.

Prova de ensaio é o conjunto de questões que deverão ser submetidas a experiência para escolha daquelas (questões) que, por apresentarem os requisitos desejados, irão figurar na prova definitiva.

Quando defrontamos o problema da construção da prova de ensaio, acodem-nos ao espírito as seguintes interrogações:

a) qual a extensão (total de questões) que deverá ter a prova de ensaio ?

b) como escolher a matéria de cada questão?

Relativamente à primeira pergunta é preciso, desde logo, considerar:

1. que a prova objetiva deve envolver uma amostra extensiva, isto é, abranger a maior área possível de conhecimentos;

2. que as questões da prova de ensaio constituirão o material donde serão selecionadas as questões da prova definitiva.

Nestas condições, é evidente que o número de questões a serem ensaiadas deverá ser o maior possível, de modo a proporcionar um material de experiência satisfatória.

Convém, entretanto, guardarmo-nos contra as limitações naturalmente impostas por certos fatores, como, por exemplo, o tempo disponível para a realização do ensaio, a natureza da matéria da prova, a fadiga dos pacientes etc.

RUCH aconselha que se inclua na prova de ensaio um número de questões 25 a 50% superior ao que se supõe necessário para a prova definitiva.

Um ponto que deve merecer a nossa atenção é o que diz respeito à distribuição do total das questões pelos assuntos da prova.

Uma boa distribuição seria, a nosso ver, aquela que reservasse 80% das questões aos conhecimentos *indispensáveis*, 15 % aos *complementares* e 5% aos *menos importantes*.

Quanto à segunda pergunta, convém, de início, salientar que a matéria da prova deverá ser:

a) *objetiva*, isto é, passível do necessário controle, de modo a eliminar qualquer dúvida ou margem de julgamento aleatório;

b) *prática*, na realidade e na aparência, isto é, em qualquer dos dois aspectos deve aproximar-se, tanto quanto possível, das condições reais de execução do trabalho que é objeto de estudo;

c) *seletiva*, isto é, distribuída em diferentes níveis de dificuldade, para poder assegurar a discriminação e classificação dos indivíduos.

Um ponto a ser considerado com especial interesse é o referente à escolha das fontes donde será extraída a matéria da prova, afim de que nesta última não sejam introduzidos:

1. os pontos de controvérsia;
2. as opiniões pessoais;
3. as hipóteses ou afirmações não confirmadas;
4. o material obsoleto;
5. os conhecimentos de nível impróprio;

6. os conhecimentos de aquisição impossível ou extremamente difícil;

7. o material, por qualquer título; impróprio ou irrelevante.

Como se compreende sem grande esforço, a escolha das fontes é de grande importância para o êxito da construção de provas. Por isso mesmo, alguns autores se tem dado o trabalho de coordenar e apreciar as fontes disponíveis ou os critérios para escolhê-las.

SYMONDS (9), por exemplo, catalogou seis critérios diferentes para a escolha da matéria de uma prova objetiva. Não obstante o trabalho deste autor se situar no campo educacional, alguns dos critérios apresentados são perfeitamente válidos para o nosso caso.

Tão importante quanto a escolha da matéria é a determinação do aspecto formal da prova. A este respeito, HAWKES, LINDQUIST e MANN observam que:

"The particular or peculiar phrasing of a test exercise, its structure, mechanical or typographical arrangement, and other external features may thus play an unsuspected but significant rôle in determining its validity". (10).

Quanto à forma, a prova de ensaio deve possuir as mesmas qualidades da prova definitiva. Não lhe devem faltar:

- a) certos requisitos de ordem geral;
- b) certos requisitos particulares.

São requisitos gerais aqueles que independem do tipo de questões adotado. Tais requisitos, em síntese, dizem respeito à clareza, correção, precisão de linguagem, ordem, estética e outros elementos indispensáveis para que as questões possam despertar interesse e ser facilmente compreendidas pelos indivíduos do nível de inteligência e de cultura exigido pelo trabalho que é objeto de estudo.

Os requisitos particulares, ao contrário, estão estreitamente ligados aos tipos de questões, variando, de certo modo, de um para outro tipo.

Vejam os quais são esses requisitos nos principais tipos de questões, a saber:

- a) completamento;
- b) múltipla escolha;
- c) formação de pares;
- d) ordenamento;

- e) certo ou errado;
- f) composição.

COMPLETAMENTO: Este tipo de questão consiste em um texto contendo claros a preencher, exemplo:

Teoricamente, boa prova é aquela cujas questões apresentam correlação com o critério e correlação entre si.

Para este tipo de questões os requisitos particulares são:

- a) o número de claros a preencher não deve ser muito grande;
- b) os claros não devem conter divisões indicativas da extensão ou número de termos a serem incluídos;
- c) os claros devem ser curtos, isto é, devem corresponder a poucas palavras, datas ou pequenas frases;
- d) as palavras que antecedem imediatamente aos claros não devem conter elemento de sugestão capaz de dificultar ou facilitar a resposta;
- e) sempre que possível, os claros deverão ser colocados na parte final, isto é, devem ser precedidos da parte já escrita da sentença;
- f) a redação, no caso de serem possíveis diversas respostas, deverá objetivar a obtenção da melhor;
- g) os claros devem corresponder a palavras, frases, etc., significativas (chaves).

MÚLTIPLA ESCOLHA: Este tipo de questão consiste numa afirmação incompleta, acompanhada de diversos termos ou frases. Destes termos ou frases apenas um ou uma (alguns, algumas) convem como complemento da afirmação exemplo:

A violência contra a coisa é dos elementos integrantes do crime de:

Sedição
Latrocínio
Furto
Rapto
Roubo
Contrabando

Para este tipo de questões os requisitos particulares são:

a) os complementos para a escolha, tanto os falsos como os verdadeiros, devem apresentar os mesmos característicos externos, isto é, os últimos não devem formar um contraste muito grande com os primeiros;

b) a relação entre o complemento e a afirmação a completar não deve ser sugerida por detalhes de redação, tais como concordância de gênero, número etc.;

c) quando houver diversos complementos verdadeiros estes não devem ser distribuídos a intervalos regulares nem seguidamente, sendo preferível intercalá-los, irregularmente, entre os complementos falsos;

d) o número de complementos falsos deve ser razoavelmente maior (5 vezes, por exemplo) do que o de complementos verdadeiros.

ORDENAMENTO. Este tipo de questões consiste na apresentação de certo número de palavras (números, partes de uma cousa, fases de uma operação, etc) a serem colocadas na devida ordem. Para este tipo de questões os requisitos particulares são: a) o número de termos (partes, fases de operações, etc.) a ordenar não deve ser muito grande; b) somente devem constar da questão termos (*idem*) que, na realidade, guardem entre si a sequência prevista; c) os termos (*idem*) apresentados, uma vez ordenados corretamente, devem formar um sentido (operações, cousa, etc.) completo; d) a possibilidade de vários arranjos verdadeiros deve ser evitada; e) em se tratando de sentenças, é inconveniente o emprego de incidências, mormente quando forem longas ou puderem ser incluídas em diferentes lugares; f) é inconveniente e por isso deve ser evitado o uso de expressões tais como: *primeiros, depois, por fim, etc.*

FORMAÇÃO DE PARES: Este tipo de questão consiste de duas relações; os termos da primeira são numerados e os da segunda, precedidos de parêntesis dentro dos quais deverá ser escrito um número ou sinal, de modo a indicar os termos que guardem entre si um elemento de associação. Exemplo:

1. Penas principais

2. Penas acessórias

- () Publicação da sentença
- () Multa
- () Suspensão dos direitos políticos
- () Detenção
- () Incapacidade para exercer função pública
- () Reclusão

Para este tipo de questão os requisitos particulares são: a) os pares a serem formados não devem apresentar-se em número excessivo nem muito reduzido (boa quantidade: 10 a 20 pares); b) a relação entre os termos não deve ser sugerida por detalhes de redação, etc; c) a diferença de assuntos não deve ser tão flagrante, de modo a sugerir sem maior esforço a relação procurada; d) quando a questão é muito extensa, as palavras (datas, etc.) a numerar devem ser arranjadas em ordem alfabética ou cronológica.

CERTO OU ERRADO: Este tipo de questão consiste em uma (ou várias) afirmação que deverá ser julgada certa ou errada, exemplo:

Para que se verifique a reincidência, é indispensável que exista uma sentença passada em julgado.

.....

Ao organizar este tipo de questões deve-se: a) evitar o emprego de duas negativas na mesma sentença; b) o emprego de afirmações parcialmente falsas, parcialmente verdadeiras; c) o emprego de períodos longos, frases com muitas incidências ou outros elementos que alterem ou introduzam solução de continuidade no desenvolvimento do raciocínio; d) a contiguidade de dois itens que guardem entre si um linha de associação, desde que a resposta de um possa sugerir a do outro.

COMPOSIÇÃO: Na sua forma mais usual este tipo de questão consiste em certo número de especificações, de acordo com as quais se deverá compor o desenho representativo de uma peça de máquina, de um movel, etc.)

Para este tipo de questão os requisitos particulares são: a) salvos os casos especiais, as especificações não devem ser muito complexas; b) quan-

do não for inconveniente a indicação de fórmulas, devem ser apresentadas todas as fórmulas possíveis ou, pelo menos, as de uso frequente; c) devem ser empregados os símbolos e convenções próprios do trabalho envolvido pela questão; d) as instruções devem ser redigidas de tal modo que a resposta possa obedecer o mais estritamente possível às regras e princípios técnicos pertinentes ao trabalho (*idem*); e) salvo os casos especiais, devem ser fornecidos os dados cuja determinação dependa de cálculos ou operações laboriosas.

PARTE III

"Fundamental principle of employment psychology.

"The tests or other measurements to be used in selecting persons for a given occupation must be evaluated by giving them to persons whose actual ability in that occupation is known and comparing efficiency in the test with efficiency in the occupation". HAROLD BURT (11).

"From the viewpoint of ultimate usefulness, one of the most important steps in the research program for the development of tests is the determination of whether the test or battery actually taps qualities of importance in success on the job, and whether it can be used in predicting such success. This involves a comparison between test score and proficiency on the job as represented in the criteria. If this comparison reveals relationship between test scores and success on the job, the tests can then be described as valid for selection purpose". — MORRIS S. VITELIS (12).

Nosso trabalho não ficou concluído com a organização da prova de ensaio. Resta-nos verificar se esta prova apresenta qualidades até aqui ainda não estudadas, mas que são indispensáveis a um bom instrumento de exame.

Dentre essas qualidades se destaca a de *constância* ou *coerência*, a qual se pode verificar com emprego:

- a) do reteste;
- b) dos grupos estatisticamente iguais;
- c) das formas equivalentes;
- d) das duas metades da prova (questões pares e questões ímpares.)

No primeiro caso (reteste), aplica-se duas ou mais vezes a prova ao mesmo grupo de indivíduos e calcula-se a correlação entre os resultados obtidos. No segundo caso, ao invés de aplicações sucessivas da prova ao mesmo grupo, ela é aplicada uma só vez a dois ou mais grupos estatisticamente iguais. Se se pretende empregar o terceiro proces-

so (formas equivalentes), constroi-se para a mesma prova um certo número (geralmente 2 ou 3) de formas paralelas, equivalentes ou semelhantes; e, uma vez aplicadas estas formas ao mesmo grupo ou a grupos estatisticamente iguais, calcula-se a correlação entre os resultados obtidos. No quarto e último caso, começa-se verificando a correlação entre os resultados da aplicação das duas metades da prova (questões pares e questões ímpares); depois, uma vez que se tem em vista utilizar a prova por inteiro, há que calcular uma segunda correlação, o que se faz com auxílio da fórmula de SPEARMAN e BROWN:

$$r_{nn} = \frac{nr}{1 + (n - 1)r}$$

Assim, aproveitando um exemplo de RUCH, o coeficiente de constância de uma prova cujas duas metades guardassem entre si uma correlação igual a 0,37, seria:

$$r_{nn} = \frac{2(0.37)}{1 + (2 - 1)(0.37)} = 0.54$$

Como deveríamos classificar a correlação encontrada (0,54)? Muito alta? Muito baixa? Não é fácil responder a esta pergunta, desde que se tenha em vista que o coeficiente de correlação não deve ser tomado como um valor absoluto. Os autores, entretanto, tem se esforçado para dar algumas indicações de ordem geral.

RUCH (13) considera baixo um coeficiente igual a 50; GARRET (14) é de opinião que o coeficiente mínimo aceitável deve ser igual a 0,80;

VITELLIS admite como satisfatório um coeficiente mínimo de 0,70.

Outra qualidade que a prova deve apresentar é a de *validade*

Quando construimos uma prova temos em vista utilizá-la para medir alguma coisa (conhecimentos, no caso vertente). Pois bem, a precisão com que a prova mede essa alguma coisa que com ela pretendemos medir é o que constitui a *validade*:

"The validity of a test or examination (ensinam JOHN A. LONG e PETER SANDFORD) (15) is usually defined as that aspect of a test which insures that it will actually measure what it claims to measure. To discover or not a test actually functions in this way, two sets of measure are obviously needed, those of the test itself and of the thing it is stacked against — the measuring rod, as it were. The latter is known as the criterion".

Se encarmos a validade do ponto de vista do trabalho de organização da prova, podemos dizer, acompanhando o pensamento de RUCH, que a validade se refere ao cuidado de incorporar à prova aqueles elementos ou itens de importância fundamental, e à preocupação de eliminar os elementos ou itens não essenciais.

Convém notar desde logo que essa *importância fundamental* deve ser compreendida não como grau de dificuldade, mas como valor funcional. Por outros termos, a importância dos elementos ou itens será tanto maior quanto mais elevado for o valor funcional dos mesmos. A esse respeito MONROE (16) ensina que:

"Under the head of validity we inquire into the degree of constancy of the functional relation existing between the scores yielded by the test and the abilities specified as being measured in the statement of its function".

A validade é geralmente interpretada em termos estatísticos por meio de coeficiente de correlação, muito embora, como veremos mais tarde, seja não somente possível como aconselhável interpretá-la por outros meios.

Para a interpretação estatística da validade necessitamos de:

a) um critério ou seja um elemento de referência com o qual serão comparados os resultados da aplicação da prova;

b) um grupo de ensaio ou seja, no nosso caso, um certo grupo de indivíduos representando diferentes graus de eficiência no trabalho que é objeto de estudo;

c) um método ou técnica de validação.

A escolha do *critério* é um dos mais delicados problemas que se apresentam no domínio da organização das provas objetivas. Isso é tanto mais compreensível se tivermos em vista que sendo o *critério* utilizado como uma variável para o cálculo do coeficiente de correlação, a significação deste último não pode deixar de ser afetada pela maior ou menor objetividade do critério.

Os critérios que tem sido objeto de preferência por parte dos técnicos são os seguintes:

1. a quantidade de produção em determinado período;

2. o julgamento dos chefes de serviço;

3. os resultados dos exames periódicos;

4. os resultados da aplicação de uma prova cujo valor como instrumento de seleção para o trabalho em estudo tenha sido devidamente verificado.

Cada um desses *critérios* apresenta, sem dúvida, seus inconvenientes.

Eis porque na prática da construção das provas objetivas se tem adotado o recurso de conjugar dois ou mais critérios.

Uma exposição detalhada dos processos usuais de combinação de diferentes critérios foi feita por HAROLD BURT (17).

Já tivemos oportunidade de dizer que os resultados da aplicação da prova de ensaio, para verificação da condição de *validade*, devem ser apreciados em relação aos graus de eficiência no trabalho.

HAROLD BURT (18), dissertando sobre a necessidade do critério, escreve:

"...Consequently, before psychological tests can be validly used for employment purposes they must themselves be tested by comparing, in a typical group of workers, efficiency in the tests with efficiency in the jobs. This implies two measures for each person on whom the tests are standardized — his test scores and some figure that represents his occupational efficiency. This latter — the thing by which the tests are actually evais technically called criterion".

Já por aí se pode avaliar a importância do *grupo de ensaio*. Infelizmente, porém, entre nós, na maioria dos casos pouco interesse se tem dispensado à organização e emprego dos *grupos de ensaio*.

Está fora de dúvida, entretanto, que é indispensável, no trabalho de validação de provas para seleção de trabalhadores, o emprego de dois processos:

a) o primeiro consistindo em aplicar a prova a trabalhadores cuja eficiência no trabalho já é conhecida; então, se os trabalhadores tidos como mais eficientes obtiverem as mais altas classificações e os trabalhadores reconhecidos como menos eficientes obtiverem as mais baixas classificações, a prova pode ser considerada válida;

b) o segundo consistindo em observar como se comportam e qual o grau de eficiência que apre-

sentam no trabalho os indivíduos selecionados pela prova.

Nenhuma prova pode ser considerada verdadeiramente válida sem que esses dois processos tenham sido empregados, ensinam MOSHER e KINGSLEY.

("No test can be said to be truly valid until both these procedures have been employed" (19).

PARTE IV

"We may desire to ascertain to what extent efficiency in a particular mental test is related to efficiency in a job. The ultimate aim is to predict one variable in terms of another; hence the need for expressing quantitatively the relation between the two variables. The correlation coefficient is the standard technique for expressing this relation". — HAROLD BURT. (20).

Dissemos, no capítulo anterior, que a validade é geralmente interpretada em termos estatísticos pelo coeficiente de correlação.

Do ponto de vista prático, devemos distinguir três hipóteses diferentes. Na primeira, o critério ou o critério e os resultados da aplicação da prova são variáveis discretas e nos informam como as classes, representadas pelas respectivas tendências centrais, se diferenciaram na realização da prova. Na segunda hipótese o critério e os resultados da aplicação da prova são variáveis contínuas e se leva em consideração, apenas, a posição acima ou abaixo na medida da tendência central em cada série. Na terceira hipótese, uma das variáveis é, ao mesmo tempo, quantitativa e contínua, enquanto a outra é dicotômica.

Na primeira hipótese, é aconselhável empregar-se o método da comparação. Para isso, divide-se o grupo de ensaio, por exemplo, em duas classes: a dos bons e a dos maus trabalhadores; traçam-se as curvas de distribuição correspondentes aos resultados da aplicação da prova nas duas classes, tendo-se o cuidado de adotar os mesmos pontos de referência.

Verifica-se a capacidade de diferenciação da prova observando o *overlapping* das duas distribuições. Quanto menor for o *overlapping*, maior será a validade da prova.

Se desejarmos obter um índice numérico da aludida capacidade de diferenciação, começaremos computando as médias e os desvios padrão dos resultados da aplicação da prova em cada classe. As fórmulas para cálculo dos desvios padrão das médias são as seguintes:

$$\sigma_m = \frac{\sigma_d}{\sqrt{N}}$$

$$\sigma_d = \sqrt{\frac{\sum f d^2}{N}}$$

Depois, calcularemos o desvio padrão da diferença entre as médias das duas classes, com emprego da fórmula apropriada:

$$f_{12} = \sqrt{\sigma_{m1} + \sigma_{m2}}$$

onde σ_{m1} e σ_{m2} são os desvios padrão das duas médias.

Determina-se o poder de diferenciação da prova comparando-se a diferença entre as médias com o desvio padrão da mesma. Se a diferença é duas ou mais vezes maior que o seu desvio padrão, a prova possui um poder de diferenciação significativo e é válida para o fim em vista.

É importante observar que o número de casos (N) não deve ser inferior a 25.

Na segunda hipótese, isto é, quando se trata de variáveis contínuas e se leva em conta, apenas, a posição acima ou abaixo da medida de tendência central das séries, o método mais adequado é o da correlação tetratórica.

O coeficiente respectivo pode ser obtido com o emprego da fórmula:

$$r = \cos \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

O método da correlação tetratórica é estudado com o necessário desenvolvimento nos bons manuais de estatística, muitos dos quais apresentam tabelas organizadas especialmente para o cálculo simplificado do valor de r .

Quando lidamos com questões que admitem duas respostas e com um critério que divide o grupo

de ensaio em duas classes, devemos dar preferência à fórmula:

$$r_{hk} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

onde a , b , c e d designam os valores inscritos em cada uma das quatro secções em que se divide o gráfico empregado, habitualmente, para representar as duas classes com seus valores acima e abaixo da média, no critério e nos resultados da aplicação da prova.

Na terceira hipótese, ou seja, quando uma das variáveis é ao mesmo tempo quantitativa e contínua, enquanto que a outra é dicotômica, o método indicado é o da correlação bi-serial. Neste caso, o cálculo do valor de r se efetua com emprego da fórmula:

$$r = \frac{(M_2 - M_1) p_q}{\sigma_z}$$

O principal inconveniente do método da correlação bi-serial reside, como observou MURILO BRAGA (21), no fato de ser o mesmo muito demorado quando temos de lidar com dados muito numerosos.

Mas em matéria de organização de provas são, justamente, os dados numerosos que asseguram resultados mais precisos; por isso mesmo se tem manifestado, da parte dos autores, uma acentuada tendência não só para simplificar os métodos de validação, senão também para torná-los comodamente aplicáveis a um número elevado de casos.

SYMONDS (22) apresenta, nesse sentido, o seu "Balance Method." Para este autor as questões que medem mais acuradamente as aptidões de uma pessoa são aquelas que ela pode resolver com 50% de precisão. Assim, a melhor prova seria aquela na qual todas as questões tivessem um grau de dificuldade tal que a média dos indivíduos do grupo pudesse respondê-las com 50% de precisão.

Um exemplo: Admitamos que tendo em vista a organização de uma prova para a seleção de telegrafistas extranumerários, submetemos o grupo de 500 indivíduos a quatro diferentes questões sobre cada um dos cinco pontos do programa de Geografia. Os dados obtidos foram os seguintes:

	Questão	Número dos indivíduos que acertaram (N_a)	Número dos indivíduos que erraram (N_e)	$N_a \times N_e$
1.º ponto	1	270	230	62.100
	2	320	180	57.600
	3	400	100	40.000
	4	250	250	62.500
2.º ponto	1	210	290	60.900
	2	280	220	61.600
	3	10	490	4.900
	4	320	180	57.600
3.º Ponto	1	250	250	62.500
	2	210	290	60.900
	3	210	290	60.900
	4	400	100	40.000
4.º ponto	1	250	250	62.500
	2	270	230	62.100
	3	400	100	40.000
	4	400	100	40.000
5.º ponto	1	320	180	57.600
	2	250	250	62.500
	3	10	490	4.900
	4	280	220	61.600

Admitindo-se que, para as questões ideais, $N_a \times N_e$ é igual a 62.500, segue-se a melhor prova que poderíamos construir seria, no caso, aquela que incluísse:

- a) para o primeiro ponto, as questões 1 e 4;
- b) para o segundo ponto, as questões 1 e 2;
- c) para o terceiro ponto, as questões 1 e 2 (ou 3);
- d) para o quarto ponto, as questões 1 e 2;
- e) para o quinto ponto, as questões 2 e 4.

CLARK (23) propõe esta fórmula extremamente simples para o cálculo da validade:

$$V = \frac{P - D}{1 - D}$$

Para ilustrar a aplicação da fórmula de CLARK, tomemos o mesmo grupo de 500 indivíduos do exemplo anterior e digamos que 200 deles erraram a questão. Assim:

$$D = \frac{200}{500} = 0,4$$

Dos 200 indivíduos que obtiveram as baixas classificações no critério, 120 erraram a questão. Assim:

$$P = \frac{120}{200} = 0,6$$

Então, substituindo-se, na fórmula, os símbolos pelos respectivos valores, teremos:

$$V = \frac{0,6 - 0,4}{1 - 0,4} = .33$$

Como se vê, a validade encontrada foi positiva, porém fraca; para que ela fosse positiva e máxima, o resultado deveria ser 1,00 ao invés de 33.

Um método de grande simplicidade e que apresenta certas vantagens, mesmo sobre a correlação bi-serial, é "LONG overlapping," segundo o qual

$$V = 1 - \frac{2 \Sigma \text{acertos abaixo dos erros}}{(\text{NP}) (\text{NE})}$$

Exemplificando: a questão *m* foi aplicada a doze indivíduos, classificados no critério (segunda coluna do quadro abaixo), pela média de produção em determinado período. Na terceira coluna do mesmo quadro estão assinalados com a letra A os indivíduos que acertaram a questão e com a letra E os indivíduos que erraram a questão:

Indivíduos	Critério	Resultados da aplicação da prova
Manoel	120	A
Pedro	110	E (4)
Joaquim	100	A
Otávio	90	A
Raimundo	80	E (2)
Luís	70	A
Orlando	60	E (1)
Osvaldo	50	E (1)
Carlos	40	E (1)
Rui	20	E (1)
Alfredo	10	A
Célio	0	E

Na linha correspondente a Pedro verificamos que, ao lado da letra *E* está escrito, entre parêntesis, o número 4; isto quer dizer que abaixo de Pedro há quatro indivíduos que acertaram a questão (Joaquim, Otávio, Luís e Alfredo). Na linha correspondente a Raimundo, vê-se, entre parêntesis, ao lado da letra *E*, o número 2, o que indica que abaixo de Raimundo há dois indivíduos que acertaram a questão (Luís e Alfredo), e assim por diante.

Aplicando a fórmula, teremos:

$$\Sigma \text{ de acertos sob erros} = 0 + 1 + 1 + 1 + 1 + 2 + 4 = 10$$

$$NA = 5$$

$$NE = 7$$

Donde :

$$V = 1 - \frac{2 (10)}{5 \times 7} = .43$$

Mas, pergunta-se, um coeficiente de correlação suficientemente alto indica com segurança a existência da validade?

TURNEY (24) responderia pela negativa, sob o fundamento de que a validade, por sua própria natureza, não pode ser determinada senão pelo julgamento dos peritos. Mais ou menos idêntico é o ponto de vista de outros autores, entre os quais MONROE (25). A maioria dos psicólogos e educadores, entretanto, aceita o coeficiente de correlação como medida da validade.

Validadas as questões não nos resta mais do que arranjá-las na prova definitiva, não esquecendo a regra geral de que nossa preferência deve recair sobre as questões que guardem alta correlação com o critério e baixa correlação entre si.

É aconselhável dispor as questões em ordem de dificuldade crescente, o que facilita extraordinariamente o trabalho de correção e ajuda a discriminar os diferentes níveis a que pertencem os indivíduos do grupo a ser examinado. Para isso, poderemos, aceitando a sugestão de EBBLEWHITE SMITH (26), dispor as questões pelo grau de dificuldade de sorte que elas se distribuam sob a forma de uma curva normal.

BIBLIOGRAFIA

- (1) *Public Personnel Administration* — Pg. 166 — Harper & Brothers — New York — 1941.
- (2) G. M. RUCH — *Objective Or New Type Examination* — Nota à pg. 9 — Scott, Foresman & Co. — New York — 1929.
- (3) *Retail Personnel Relations* — Pg. 43 — Prentice Hall Co. — New York — 1939.
- (4) *Job Analysis and The Curriculum* — Pg. 22 — Williams & Wilkins Co. — Baltimore — 1923.
- (5) HERBERT MOORE — *Psychology For Business And Industry* — Pg. 176 — McGraw Hill Co. — New York — 1939.
- (6) *La Psychotechnique* — Pg. 70 — Comité Central Industriel de Belgique.
- (7) *Les Examens d'Aptitude Professionnelle* — Trad. de MARCEL THIERS — Pg. 69 — Dunod — Paris.

-
- (8) *The Construction And Use Of Achievement Examinations* — Pg. 5 — Houghton Miffling Co. — New York — 1936.
- (9) *Measurement in Secondary Education* — Macmillan Co. — New York — 1934.
- (10) *Op. cit.* — Pg. 79.
- (11) *Principles of Employment Psychology* — Pg. 5 — Harper & Brothers — New York — 1942.
- (12) *Industrial Psychology* — Pg. 249 — W. W. Norton & Co. New York — 1932.
- (13) *Op. cit.* — Pg. 90.
- (14) *Statistics in Psychology and Education* — Pg. 269 — New York — 1926.
- (15) *The Validation of Test Items* — Pg. 6 — The Department of Educational Research-University of Toronto — 1935.
- (16) *An Introduction to The Theory of Educational Measurement* — Houghton Miffling Co. — Boston — 1923.
- (17) *Op. cit.* — Pg. 188 e sgs.
- (18) *Op. cit.* — Pg. 171.
- (19) *Op. cit.* — Pg. 174.
- (20) *Op. cit.* — Pg. 150.
- (21) *Validade e Fidedignidade nos testes coletivos de inteligência* — pg. 28 — Rio — 1938.
- (22) *Choice of items for a test on the basis of difficulty* — "Journal Ed. Psy." — New York — Outubro de 1929.
- (23) *A Method of evaluating the units of a test* — Idem — Abril de 1928.
- (24) *The concept of validity in mental and achievement testing* — "Journal of Education Psychology" — New York — Fevereiro de 1934.
- (25) *Hazards in the measurement of achievement* — "School and Society" — New York — 1935.
- (26) *The Construction and Validation of a Group Test of Intelligence Using The Spearman Technique* — "The Educational Research" — Ontario College Education — 1935.
-