Observações sôbre os testes de múltipla escolha

CRISTOVAM DIAS GASPAR
Assistente de Seleção do D.A.S.P.

QUANDO se fala em organizar provas de conhecimento para a seleção de pessoal, surgem imediatamente dois problemas fundamentais: como elaborar as questões e como empregá-las de maneira que se possam selecionar e classificar, ao mesmo tempo, os elementos do grupo.

Antes de tratarmos dos testes de múltipla escolha, desejamos salientar algumas condições, às quais um examinador deve estar sujeito, tendo-se em vista que:

a) *um domínio seguro* da matéria se harmoniza com o poder de criação, de invenção e de originalidade. Quem não conhece bem a disciplina utiliza situações conhecidas, copia formas de compêndios ou propõe problemas clássicos. Sua incapacidade em classificar questões para os diversos níveis não lhe permitirá a organização de formas paralelas. O senso prático das dificuldades é uma função do referido domínio. Não seria interessante treinar um indivíduo que nem ao menos tivesse a noção clara do grupamento bruto das questões em graus de dificuldade;

b) *a experiência em classe* é um dos melhores fatores de êxito, não só por requerer aptidões condizentes com a função, como também por apresentar idênticas fases de treinamento no trabalho. Ensinar e examinar estão intimamente ligados;

c) *o método de organização, a vontade de produzir, a paciência e o zêlo* são elementos indispensáveis para a organização de provas padronizadas. Cumpre ao examinador adquirir a sensibilidade necessária, afim de que possa lidar com todas as medidas, calibrar os instrumentos, notar variações, anomalias e defeitos que porventura se apresentem. E' necessário muito amor ao trabalho; não se deve deixar de fazer os retoques, de mudar, medir e mesmo experimentar, sempre que se tornar conveniente.

A APRESENTAÇÃO DAS QUESTÕES

Na organização dos testes de múltipla escolha o examinador emprega as melhores manifestações de seu pensamento, traduzidas por uma apresentação de questões acompanhadas de formas despistatórias para o julgamento objetivo, considerando que o raciocínio observado nas soluções não deve ser acompanhado de rascunhos ou cálculos escritos. Êsse tipo talvez seja o mais difícil de organizar, dada a grande oscilação do grau de dificuldade que o mesmo apresenta. Não é raro um examinador atrapalhar-se com a constância de suas questões e surpreender-se com o deslocamento do grau de dificuldade relativa motivado por leves modificações nas formas paralelas ou nos elementos das alternativas.

Admitamos que se queira verificar o conhecimento das operações com frações decimais, e se proponha a seguinte unidade:

"*Sublinhe o resultado correto do seguinte cociente*:

0,04÷0,002 = (0,002); (0,2); (20); (200) ou (0,02)?

Aquí não se trata de verificar se o candidato sabe dividir o número *4* pelo número *2*. A divisão de inteiro por inteiro poderia ser testada por meio de uma outra questão. O mais importante é a colocação da vírgula decimal.

Vejamos as fases do raciocínio empregado na resolução:

a) o candidato considera 0,04 e 0,002 como números inteiros e para isso multiplica o primeiro por 100 e o segundo por 1.000;

b) observa que o cociente assim obtido vem dividido por 10. Torna-se necessário multiplicá-lo por *10*, afim de obter o verdadeiro cociente;

c) de posse dêsse dado (20), o candidato marcará a terceira resposta.

Entretanto, se tivéssemos a seguinte apresentação:

0,04÷0,002 = (0,4÷2); (4÷2); (40÷2); (4÷0,02) ou (0,02), a dificuldade já não seria a mesma. A experiência tem revelado que o deslocamento do "standard score" é aproximadamente de + 0,2.

Se, por outro lado, a resposta certa estivesse misturada com absurdos visìvelmente grosseiros, o "standard score" caïria por completo e a questão teria valor nulo para a seleção.

Como exemplo, vejamos o caso anterior sob outro aspecto:

0,04÷0,002 = (2×4); (0,04); (0,002); (20) ou (1943)?

O candidato não teria necessidade do conhecimento que se pretendeu apurar, para apontar a resposta certa.

Um caso interessante pode ser observado, quando o examinador calcula o tempo de resolução para a bateria de testes de múltipla escolha com duas respostas certas, em cada unidade, baseado no tipo de um êrro único. Verifica-se, experimentalmente, que cêrca de 50% dos candidatos alcançam sòmente 3/5 da prova. O estudo estatístico da distribuïção mostra claramente a anomalia, bastando, para isso, que se escale a mesma em têrmos do desvio-padrão. Na hipótese de mais de uma resposta para o próprio texto da pergunta formulada, cuja solução faz parte de cinco alternativas compostas, aparece outro fator da probabilidade de acertos por acaso que, entretanto, aumenta a dificuldade, exigindo maior tempo de resolução. Suponhamos, para maior clareza, que se tenha dado um texto A, contendo três erros, respectivamente, de ortografia (*g*), de concordância (*e*) e de regência (*r*). Apresentando sob a forma de múltipla escolha, consideremos seis alternativas, por exemplo, B, C, D, E, F e G.

alternativa — B — contém o êrro *g*
" — C — contém o êrro *e*
" — D — contém o êrro *r*
" — E — não contém êrro algum
" — F — contém os êrros *e* e *r*
" — G — contém os êrros *g* e *r*.

Pode-se observar que êste teste representa um grupo de sete questões de textos simples. Aquí o candidato sujeita todos os itens, A, B, C, D, E, F e G à correção.

Sòmente a apuração e validação dêsses itens vem demonstrar aquilo que muitos não percebem imediatamente.

Outra modalidade do teste de múltipla escolha consiste na obrigação de assinalar duas ou três respostas certas dentre cinco ou seis, de modo que qualquer êrro anule a questão. Como exemplo típico citaremos a seguinte:

"Marque com uma cruz (+) os resultados correspondentes ao valor do produto 3x8".

| ( ) 12x3 | ( ) 12x2 | ( ) 6x4 | ( ) 5x6 |
| ( ) 21x1 | ( ) 18+6|

Basta que o candidato assinale um resultado errado para ficar provado que desconhece o verdadeiro valor do produto. Um conjunto bem organizado dêsse tipo apresenta uma probabilidade muito pequena de acertos por acaso. Para uma unidade temos uma probabilidade de 1 para 120; para um grupo de 5 questões uma probabilidade ainda menor: 1 contra 2.483.200.000 casos.

A INFLUÊNCIA DO ACASO

Outras questões assás interessantes e que prendem todo o assunto dos testes de várias alternativas são: a interpretação dos resultados, a ponderação dos itens e a renovação das formas, utilizando-se em grande parte as noções elementares do cálculo de probabilidades. Muitos não acreditam na influência do acaso e assim quando tratam dêsses testes, não se referem às correções nem às compensações tão necessárias. ASKER, escrevendo um artigo intitulado "The Reliability of Tests Requiring Alternatives Responses" (1), chama a atenção dos estudiosos para o problema.

Organizemos uma tabela e calculemos as probabilidades de todas as situações, em *n* questões de duas alternativas. Chamando *f* a freqüência das respostas falsas e *v* a das verdadeiras, desenvolvendo

$$(v+f)^n = \sum_{i=0}^{n} \binom{n}{i} f^i v^{n-i}$$

obteremos o têrmo genérico de ordem

$$(i+1):\ T_{i+1} = \binom{n}{i} f^i v^{n-i}$$

(1) *Journal of Educational Research* — 1924 — Vol. IX.

com uma situação de i respostas falsas e (n-1) verdadeiras e cuja probabilidade de ocorrência é:

$$\binom{n}{i} 2^{-n}$$

Pelo desenvolvimento podem-se observar as probabilidades das situações:

SITUAÇÕES	PROBAB.
n verdadeiras e *o* falsas:	$\binom{n}{0} 2^{-n}$
(n-1) verdadeiras e 1 falsas:	$\binom{n}{1} 2^{-n}$
(n-2) » e 2 »	$\binom{n}{2} 2^{-n}$
. . .	.
. . .	.
. . .	.
2 » (n-2) »	$\binom{n}{2} 2^{-n}$
1 » (n-1) »	$\binom{n}{1} 2^{-n}$
0 » n »	$\binom{n}{0} 2^{-n}$

Para n=20, teremos o exemplo de RUCH (2)

V	f	$\binom{n}{i}$	Prob. 1 para:
20	0	1	1.048.576
19	1	20	52.429
18	2	190	5.519
17	3	1.140	920
16	4	4.845	216
15	5	15.504	68
14	6	38.760	27
13	7	77.520	13,5
12	8	125.970	8,3
11	9	184.756	5,7
10	10	167.960	6,2
..	..	....	...
..	..	....	...
3	17	1.140	920
2	18	190	5.519
1	19	20	52.429
0	20	1	1.048.576

AS FÓRMULAS EMPÍRICAS

Podemos então estabelecer o seguinte raciocínio para a determinação de fórmulas corretoras:

Consideremos uma prova com N questões de duas escolhas. Suponhamos que um candidato conheça P % da referida prova, isto é, que seja capaz de responder corretamente a P % das questões. A parte restante (100-P) % só será acertada por acaso.

Para um número suficientemente grande de observações seremos levados a considerar o valor mais provável da percentagem dos acertos por acaso como $\frac{(100-P)\ \%}{2}$. O candidato poderá apresentar, então, uma percentagem total de acertos igual a $\frac{(100+P)\ \%}{2}$. Como devemos eliminar a influência do acaso, isto é, não permitir que as questões acertadas por sorte modifiquem a nota verdadeiramente merecida ou a nota correspondente aos acertos concientes, poderemos subtrair $\frac{(100-P)\ \%}{2}$ de $\frac{(100+P)\ \%}{2}$, obtendo, assim, o valor real das percentagens.

Dêsse modo, poderemos compensar os resultados provenientes da influência do acaso, conforme se segue:

% verdadeira de acertos = % total de acertos — % de êrros.

Daí a fórmula:

$$A = A_t - E \quad \begin{cases} A_t = \text{acertos totais} \\ E = \text{erros totais} \end{cases}$$

Caso N não seja suficientemente grande, é bem possível que um candidato conhecedor sòmente de uma pequena percentagem das questões fique em igualdade de condições com um outro que conheça uma grande percentagem. Pràticamente é impossível obtermos uma prova com um número de questões suficientemente grande para a segurança dêste processo, por exemplo, com 3.000 a 5.000 questões. Cumpre, então, diminuir a probabilidade dos acertos por acaso, aumentando o número de casos possíveis nas alternativas de cada item.

Suponhamos que uma prova apresente N questões de n > 2 alternativas. A probabilidade de acertos por acaso será $p = \frac{1}{n}$ e a contrária $q = \frac{n-1}{n}$. Os valores mais prováveis serão res-

(2) RUCH G. N., *The Objective or New-Type Examination*, pág. 330-331.

pectivamente: $\frac{N}{n}$ e $\frac{N}{n}(n-1)$.

Como o valor dos acertos por acaso (A_a) não vem explícito, pois conhecemos sòmente o valor dos acertos totais (A_T) e o dos êrros (E), poderemos calcular o valor de (A_a) em função de (E).

Sendo $\frac{N}{n}(n-1)$ o valor mais provável de E, faremos:

$$\frac{N}{n} = \frac{E}{n-1}$$

$\frac{N}{n}$ é o valor esperado de (A_a), logo:

$$A_a = A - \frac{E}{n-1}$$

No cálculo do "standard score" compensado de cada item, poderemos proceder do mesmo modo, substituindo as percentagens brutas por percentagens ajustadas. Se, nos exemplos, tivermos uma bateria de testes com *n* alternativas, apresentando a seguinte distribuïção:

N. das Questões	Freqüência dos acertos	Freqüência relativa dos acertos
1.º	f_1	r_1
2.º	f_2	r_2
3.º	f_3	"
"	"	"
"	"	"
i	f_i	r_i
"	"	"

Poderemos considerar a freqüência genérica r_i a dos acertos e $1 - r_i$ a dos êrros, cujas probabilidades $\frac{1}{n}$ e $\frac{n-1}{n}$, respectivamente, nos fornecerão a relação $\frac{1-r_i}{n-1}$ como o valor mais provável de acêrto por acaso. A freqüência correta será:

$$r_c = r_i - \frac{1-r_i}{n-1}$$

ou em têrmos de percentagens:

$$P_c = P_i - \frac{100-P_i}{n-1}$$

P_c = percentagem correta

P_i = percentagem de acertos observada

n = número de alternativas.

Para maior facilidade de aplicação pode-se organizar uma tabela de dupla entrada e calcular ràpidamente as correções necessárias a introduzir no "standard-score" bruto. Nota-se, então, que as percentagens corretas aumentam à medida que *n* cresce. Não é difícil provar, entretanto, que:

$$P_c = \lim_{n=\infty} \left[P_i - \frac{100 - P_i}{n-1} \right] = P_i$$

A percentagem observada só coincidirá com a percentagem ajustada quando se tiver um número infinitamente grande de alternativas.

Combinando os dois casos, isto é, o número de questões com o número de escolhas, será possível obter-se um elevado grau de segurança para êsses tipos de questões.